THE UNIVERSITY
of EDINBURGH

The human neuronal LUHMES cell line as a model system for studying Rett syndrome

Ruth R. Shah

June 2017

Thesis submitted for the degree of Doctor of Philosophy
Wellcome Trust Centre for Cell Biology
University of Edinburgh

Declaration

I declare that this thesis has been composed by myself and the work presented herein is my own, except where stated otherwise. This research has not been submitted for any other degree except as specified.

Ruth R. Shah

Edinburgh

June, 2017

Acknowledgements

First and foremost, I would like to thank my supervisor, Prof Adrian Bird for his support and encouragement throughout the 4 years of this PhD. As the focus of my project shifted over the years he helped me find my path, always had new ideas and experiments for me to consider, and I am grateful for the scientific freedom that he gave me with this project. With such a broad range of experiments and no strict hypothesis testing, I often found myself getting lost in the details, but Adrian always managed to shine a light on the bigger picture and helped me to see things in the correct context.

Secondly, thanks have to go to Dr Justyna Cholewa-Waclaw. Over the past 4 years we have been working together to establish LUHMES cells as model system for Rett syndrome. It has been a difficult journey but neither this project nor our joint endeavours would have been possible without the continuous discussion, debate and collaboration that she has offered me. I want to thank her for always being committed to the project and for persevering when things were difficult. I look forward to the final chapter of our story as we bring this project to publication and I cannot wait to celebrate with you!

There are numerous other people that have contributed in one way or another towards the completion of this thesis and while I am sure I will not mention everyone I would like to at least try. Big thanks go to Kashyap Chaatbar for making endless Venn diagrams for me! He is my bioinformatics guru and never batted an eyelid when colours needed changing, different significance cut-offs were required or when never-ending line graphs were needed. Christine Struthers for keeping the lab in one piece and for always having the answer to "Christiiine…" Jim Selfridge for being the best bench buddy anyone could ask for. You always ask the most pertinent questions about my project, are always willing to help out in any way possible, and truly have the best life advice. Remember; keep an eye out for that

pesky squirrel. Becky Tillotson for being my PhD buddy and for telling me everything I needed to know about writing and submitting this thesis. Also for persuading me to take a day off thesis writing for my birthday.. and for providing the prosecco! Mike Robson for teaching me the process of Nucleofection, without which this LUHMES project would have faltered at the first hurdle. Monica Coenraads and the folk at RSRT for organising such brilliant MeCP2 consortium meeting every year. These meetings have allowed me to share my data in a productive and stimulating environment and because of these meetings I now have a short-term collaborative LUHMES project in Portland, Oregon to look forward to. Katie Paton for being a hard-working and extremely productive student and Laura Fitz-Patrick for taking over the LUHMES baton from me. I wish you both endless successes in your PhDs and advise you both to never take it too seriously!

Finally, of course, I have to thank the Wellcome Trust for financing this 4-year studentship and the Wellcome Trust Centre for Cell Biology for selecting me as a PhD candidate. Without these two areas of support I would not be where I am today.

Abstract

Rett syndrome (RTT) is a severe neurological disorder that affects approximately 1:10000 girls. Classical RTT is defined by a developmental regression phase and subsequent stabilisation of diagnostic criteria, which include partial or complete loss of spoken language, dyspraxic gait and stereotypic hand movements such as hand mouthing. RTT is a monogenic disorder, with the majority of cases being due to loss-of-function mutations in MeCP2 (methyl-CpG binding protein 2). Due to this clear genotype-phenotype link multiple RTT mouse models have been used to elucidate the molecular details, and consequent neuropathogenesis, of this complex neurological disease, as well as for the development of potential therapeutics for RTT. However, as the molecular details become clearer, the need for a simpler model system becomes evident. Human induced pluripotent stem cells (hiPSCs) generated from RTT patient fibroblasts are an option; however the handling of these cells is laborious, time-consuming and expensive and they often differentiate into a heterogeneous population of cells. To explore an alternative human model system I have been genetically engineering and experimenting with the human dopaminergic LUHMES cell line. LUHMES cells are an immortalised pre-neuronal cell line derived from an 8-week old, female foetus and can readily be differentiated into a homogeneous population of mature, electrically active neurons in just one week. In this thesis I have assessed the phenotypic properties of the wild-type cell line, demonstrated the ease of genetic manipulation of LUHMES cells by CRISPR/Cas9 approaches, generated seven mutant *MECP2* LUHMES cell lines and explored the potential of protein therapy as a therapeutic approach for RTT. The LUHMES cell line proves to be extremely easy to handle and robust and has yielded novel molecular insights into the function of MeCP2 in human neurons. In particular, MeCP2-null cells show a striking relationship between the level of gene body methylation and the extent of transcriptional upregulation when compared to wild-type neurons. In contrast neurons that express a form of MeCP2 that can bind to DNA but cannot

recruit a transcriptional corepressor complex (the R306C mutant) do not exhibit substantial gene expression alterations, yet do display a consistent decrease in total RNA amount. This decrease in total RNA is recapitulated in MeCP2-null LUHMES-derived neurons and in brain regions from MeCP2-R306C mice. The requirement for functional DNA binding for normal gene-body methylation dependent gene repression is demonstrated by assessing LUHMES cells that overexpress MeCP2-R111G, a protein that cannot bind to DNA. Furthermore, overexpression of the MeCP2-R306C protein highlights the importance of NCoR binding for normal gene repression, but also demonstrates that MeCP2-R306C protein retains some gene repression activity. Thinking more broadly, this cell line also has applications as a model system for a variety of other neurological disorders; as a simplified model system to elucidate molecular and neurological phenotypes, and as a relevant human system that can be cultured in a high-throughput manner for testing therapeutic strategies.

Lay Summary

Rett syndrome is a rare neurological disorder that has a frequency of 1 in 10000 live female births. It is a severe disorder characterised by apparent normal development for the first 12-18 months of life, followed by a loss of motor skills, loss of acquired speech, stereotypical hand movements, seizures, cardiac problems and hyperventilation. Unlike other autism-associated disorders, Rett syndrome is predominantly caused by mutations in a single gene, *MECP2*. Rett syndrome is also exceptional due to the discovery that in mice, the majority of phenotypes can be reversed if a functional copy of *MeCP2* is restored. With the knowledge that Rett syndrome is potentially curable, the field is now working towards developing therapeutics to treat, and hopefully cure the disorder. One avenue towards achieving such a goal is to understand the function of the MeCP2 protein in cells.

MeCP2, or methyl-CpG binding protein 2, is a DNA-binding protein that is highly expressed in the brain, particularly in neurons. MeCP2 recognises a DNA modification called a methylated-CpG dinucleotide, or mCpG. The mCpG mark is associated with the repression of gene expression and thus it is assumed that MeCP2 binds to this modification and contributes to gene repression in neurons. This suggests that upon mutation of MeCP2, this repressive activity is lost, and the resulting perturbation of gene expression leads to the Rett syndrome phenotype. As such, one would expect that upon deletion or mutation of MeCP2, the expression level of a number of genes would be increased. In contrast, gene expression both increases and decreases thus complicating the interpretation of the function of MeCP2.

Most studies have used mouse models of Rett syndrome, which though valuable, have limitations. The rodent brain contains multiple different cell types that may or may not accurately reflect the function of MeCP2 in humans. This thesis, therefore, will assess the potential for a human, neuronal cell line called LUHMES cells as a new model system for

studying MeCP2 and Rett syndrome. Firstly, I establish that LUHMES cells are a robust, easy to handle cell line that can be rapidly differentiated into mature, electrically active neurons and can be genetically manipulated to create novel MeCP2 mutant cell lines. Secondly, I demonstrate that MeCP2 mutant LUHMES cells can differentiate into a population of mature neurons, thus expelling the hypothesis that MeCP2 mutations inevitably cause neurodevelopmental defects. Thirdly, assessment of RNA levels (the product of gene expression) in MeCP2 mutant neurons reveals potential mechanisms for how MeCP2 regulates gene expression, as well as revealing novel defects in total levels of RNA. Finally, this thesis also assesses the use of LUHMES cells as a tool for developing therapeutic strategies by demonstrating that MeCP2 protein itself can enter mature, human neurons and thus providing preliminary evidence for "protein therapy" as an avenue for Rett syndrome treatment.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Full title
ASD	Autism spectrum disorder
cDNA	Complementary RNA
CA	Cytosine + adenosine dinucleotide
CD	Circular dichroism
CH	Cytosine followed by any bas other than guanine
CG	Cytosine + guanine dinucleotide
Co-IP	Co-immunoprecipitation
CRISPR	Clustered regularly interspaced short palindromic repeats
DAT	Dopamine transporter
DNA	Deoxyribonucleic acid
DSB	Double strand break
FC	Fold-change
gDNA	Genomic DNA
gRNA	Guide RNA
HDAC3	Histone deactylase 3
HDR	Homology directed repair
hESCs	Human embryonic stem cells
hiPSCs	Human induced pluripotent stem cells
HPLC	High performance liquid chromatography
KI	Knock-in
KO	Knock-out
LUHMES	Lund human mesencephalic
MAP2	Microtubule associated protein 2
MBD	Methylated DNA binding domain
mCpA	CA dinucleotide containing a methylated cytosine residue
mCpG	CG dinucleotide containing a methylated cytosine residue
MeCP2	Methylated CpG-binding protein 2
MEFs	Mouse embryonic fibroblasts
mESCs	Mouse embryonic stem cells
miPSCs	Mouse induced pluripotent stem cells
MOI	Multiplicity of infection
mRNA	Messenger RNA
NCoR (aka NCoR1)	Nuclear receptor corepressor
NF	Neurofilament
NID	NCoR interaction domain
NLS	Nuclear localisation signal

NMR	Nuclear magnetic resonance
ORF	Open reading frame
PSD95	Post-synaptic density protein 95
RNA	Ribonucleic acid
mRNA	Messenger RNA
rRNA	Ribosomal RNA
RTT	Rett syndrome
sgRNA	Small guide RNA molecule
SMRT (aka NCoR2)	Silencing mediator for retinoid and thyroid receptors
ssODN	Single-stranded oligodeoxynucleotide
TBL1 (aka TBL1X)	Transducin Beta Like 1 X-Linked Receptor 1
TBL1R1 (aka TBL1XR1)	Transducin Beta Like 1 X-Linked Receptor 1
TH	Tyrosine hydroxylase
TRD	Transcription repression domain
TuJ	Neuronal-specific βIII-tubulin
XCI	X chromosome inactivation

Chapter 1

1. Introduction

1.1 Vertebrate DNA methylation

Every cell in our body contains the same complement of genes and DNA sequences, so the question of how, for example, a skin cell is distinguished from a liver cell is an important question. One answer lies with the regulation of gene expression. While every cell in our body contains the same set of genes, only certain subsets of these are expressed in each tissue and cell type. There are a number of mechanisms that contribute to tissue-specific gene regulation including the modification of histone tails, organisation in 3-dimensional space, tissue-specific and DNA-sequence specific transcription factors, and DNA methylation. DNA methylation is a true epigenetic mechanism as its properties can be passed from mother to daughter cell. It is found in a variety of species including bacteria, invertebrates, plants, mammals and fish, although the distribution of DNA methylation varies from organism to organism (Hendrich & Tweedie, 2003). The importance of this epigenetic modification is underscored by its requirement for proper embryonic development (Li *et al.*, 1992), its disruption as a hallmark for many diseases including cancer (Schubeler, 2015), and the presence of mutations in proteins that recognise the methylation mark being associated with neurological disorders (Amir *et al.*, 1999).

1.1.1 An introduction to DNA methylation

DNA methylation was first discovered in 1948 by paper chromatography experiments that separated individual nucleoside molecules from calf thymus samples. The separation of deoxyribonucleic acid samples revealed a peak that corresponded to an unknown substance (Hotchkiss, 1948). This peak had similar characteristics to cytosine, and its relationship to

cytosine was similar to the relationship of thymine (5-methyluracil) to uracil. Thus it was hypothesised that this so-called "epicytosine" mark was in fact 5-methylcytosine (Hotchkiss, 1948). Since then, the 5-methylcytosine (m^5C) mark has been found in multiple organisms in the context of a symmetrical CpG dinucleotide, allowing for the preservation of the DNA modification during cell division via semi-conservative DNA replication (Bird, 1978; Bird & Taggert, 1980; Holliday & Pugh, 1975; Riggs, 1975).

Vertebrate genomes are globally methylated at a level of 60-90%, with the exception of so-called CpG islands (CGIs). CGIs are ~ 1kb-long stretches of GC-rich, CpG-rich DNA that are unmethylated and often found associated with gene promoters. Developmentally regulated methylation of CGIs, though relatively rare, is associated with silencing of the associated gene (Bird, 1986; Illingworth *et al.*, 2010; Tweedie *et al.*, 1997). The remaining unmethylated CGIs that are not associated with a known gene promoter (e.g. intragenic CGIs) are often associated with active transcription and H3K4me3, suggesting a potential role for intragenic CGIs in regulating alternative promoters that are situated within gene bodies (Maunakea *et al.*, 2010). In the vertebrate genome, the CpG dinucleotide, while being frequently methylated, is found at frequency lower than expected (~ 1/5) (Bird *et al.*, 1985; Bird, 1986). The reason for this unusual abundance of mCpG is due to the mutability of the mC mark. This base has a tendency to be mutated to thymine, thus over evolutionary time mC can be converted to T resulting in a lower than expected CpG frequency in the bulk genome. CGIs, however, remain unmethylated and therefore are not subjected to mCp mutation pressures and thus retain the expected frequency of CpGs for their CG densities. It is thought that the active binding of transcription factors at these promoter-proximal sites prevents the methylation-depositing machinery from gaining access and thus contributes to the methylation-free, CpG-high status of CGIs (Bird, 1986).

The methyl modification is placed on position 5 in the cytosine ring by DNA methyltransferase enzymes, or DNMTs, using S-adenosyl-L-methionine (SAM) as a methyl donor group. DNMT3a and DNMT3b are known as the *de novo* methyltransferases as these two enzymes have the ability to place a methyl group on a cytosine ring, in the absence of any pre-existing methylation (Okano *et al.*, 1999). DNMT1 is known as the maintenance DNMT due to its ability to recognise hemi-methylated DNA (for example, immediately after DNA replication) and place a methyl group on the un-modified cytosine (Jones & Liang, 2009). The importance of DNMTs is underscored in mouse models where knock-outs (KOs) for the DNMTs are embryonic lethal (Li *et al.*, 1992; Okano *et al.*, 1999), and in human

patients where mutations within them are frequently associated with diseases such as cancer (Jones & Liang, 2009).

1.1.2 Methylated DNA binding proteins

The discovery of proteins that can read and write the mCpG DNA modification was crucial to establishing the importance of this mark. The first protein that was discovered to bind to methylated DNA was eukaryotic MDBP (methylated DNA-binding protein) (Huang *et al.*, 1984). Further analysis of the binding specificities of this protein *in vivo* in human placenta discovered that in fact MDBP could bind to non-methylated sites, due to the replacement of methyl-cytosine with thymine (which contains a methyl group, also on the carbon-5 position of the pyrimidine ring) (Zhang *et al.*, 1989). MeCP1 (methyl-CpG binding protein 1) was therefore the first identified protein that could bind specifically to DNA only when it was methylated (Meehan *et al.*, 1989). MeCP1 requires at least 12 mCpG dinucleotides for binding and can bring about gene repression in both *in vitro* and *in vivo* studies (Boyes & Bird, 1991; Meehan *et al.*, 1989). A 2nd protein with methylation-specific DNA binding properties was reported three years after the discovery of MeCP1, and was termed MeCP2 (Lewis *et al.*, 1992). Like MeCP1, MeCP2 could repress transcription both *in vitro* and *in vivo*, but unlike MeCP1, MeCP2 required only a single mCpG dinucleotide for binding and repression (Lewis *et al.*, 1992; Meehan *et al.*, 1992). A 3rd methylated DNA binding domain (MBD)-containing protein was later identified called MBD1 (originally called PCM1 for protein containing MBD 1) that, despite having its own transcriptional repression domain (TRD), may be part of a complex with MeCP1 and also contains CXXC domains which can bind to non-methylated DNA (Cross *et al.*, 1997; Ng *et al.*, 2000). The CXXC and MBD domains of MBD1 mean that MBD1 can be recruited by both methylated and non-methylated CpG dinucleotides, a unique property that means MBD1 can exert a repressive effect upon transcription regardless of the methylation status of the DNA (Fujita *et al.*, 1999; Jorgensen *et al.*, 2004; Ng *et al.*, 2000).

With the sequence of the methylated-DNA binding domain (MBD) of MeCP2 to hand (Lewis *et al.*, 1992), database searches identified a whole family of proteins that also contain an MBD-like domain; MBD1, MBD2a, MBD2b, MBD3, MBD4, MeCP1 and MeCP2 (Hendrich & Bird, 1998). MBD2 and MBD3 have 71.1% amino acid sequence similarity, but apart from that the only sequence homology between the members of the family is the MBD itself. All proteins are expressed in a variety of somatic tissues, are absent or expressed at low levels in embryonic stem cells (ESCs) and can bind to methylated DNA *in vitro*,

although the DNA binding property of MBD3 is not specific to methylated DNA. Consistent with this, MBD-1, -2 and -4 all localise to mCpG-rich heterochromatic foci in mouse cells, whereas MBD3 demonstrates diffuse nuclear staining or occasional foci that do not co-localise with heterochromatic foci (Hendrich & Bird, 1998).

Two other MBD-containing proteins (MBD5 and MBD6) were identified using less stringent bioinformatics approaches concerning homology within the MBD domain (Laget *et al.*, 2010; Roloff *et al.*, 2003). Despite localising to heterochromatic foci in cells, MBD5 and MBD6 cannot bind to methylated DNA *in vitro* and do not lose cellular localisation upon DNA methylation depletion, thus casting doubts over their roles as methylation-specific DNA binders (Laget *et al.*, 2010). Finally, a family of proteins (the Kaiso family) that do not contain a canonical MBD but can bind to methylated DNA have also been identified (Filion *et al.*, 2006; Prokhortchouk *et al.*, 2002). Kaiso in particular is known to interact with NCoR (a component of the NCoR/SMRT transcriptional repressor complex) and lead to decreased transcription *in vivo* (Yoon *et al.*, 2003b).

In terms of expression profile, all of the methylated DNA binding proteins studied have a broad expression pattern throughout the body, but in contrast to the mark that they bind, are not required for embryogenesis suggesting potential redundancy between certain MBD proteins. In order to identify if there was functional redundancy, MeCP2 null mice were crossed with MBD2 null mice to create double KO mice. The phenotypes of these mice were no different to that of MeCP2 null mice and thus there appears to be no functional redundancy of these two proteins for development (Guy *et al.*, 2001). Triple KO of MeCP2, MBD2 and Kaiso also demonstrated no altered phenotype compared to MeCP2 null mice (Caballero *et al.*, 2009). This study also showed that the triple KO neural stem cells retained the full potential to differentiate into astrocytes and neurons, but differentiation into neurons was somewhat delayed as determined by expression of βIII-tubulin. This reduced expression level, however, was not found in either of the single KO cells thus suggesting some redundancy between at least 2 of these 3 proteins (Caballero *et al.*, 2009).

Thus to conclude, the MBD-containing MeCPs include MBD1, MBD2, MBD4 and MeCP2, while Kaiso and its family of proteins bind to methylated DNA using zinc-finger domains. None of these proteins are essential for embryonic development and MeCP2 null mice appear to exhibit the most severe phenotype. While in transfection experiments these

proteins can all bind to and repress the same reporter gene, there appears to be no functional redundancy between these proteins *in vivo*.

1.1.3 DNA methylation as a transcriptional repressive mark

DNA methylation is thought to be involved in gene repression and there is much evidence in support of this. Cells that express the chicken β-actin gene contain unmethylated sites at the end of the gene, whereas cells that do not express this gene are generally methylated at this site (McGhee & Ginder, 1979). Cytidine analogues such as 5-aza-cytidine that inhibit DNA methylation result in differentiation phenotypes (Jones & Taylor, 1980; Taylor & Jones, 1979) and cause the upregulation of genes on the inactive X chromosome (Mohandas *et al.*, 1981). Furthermore, methylation of reporter genes prior to transfecting them into cells results in the absence of gene expression of that reporter gene (Boyes & Bird, 1991; Buschhausen *et al.*, 1985; Busslinger *et al.*, 1983). More recent genome-wide studies looking at the correlations between DNA methylation and endogenous transcription levels also confirm the association of DNA methylation with a repressive environment, at least at CGI promoters (Illingworth *et al.*, 2008; Weber *et al.*, 2007).

Potential mechanisms for how DNA methylation may regulate transcription include the direct inhibition of transcription factor binding and indirect mechanisms involving recruitment of MBD-containing proteins. Evidence of the former includes the inability of the transcription factor AP-2 to bind to the proenkephalin gene upon CpG methylation (Comb & M.Goodman, 1990) and a loss of transcriptional activity upon methylation of the cAMP-responsive element in PC12 and HeLa cells (Iguchi-Ariga & Schaffner, 1989). Nevertheless the observation that methylation of the Herpes simplex virus thymidine kinase promoter reduces transcriptional output from the promoter, but does not affect that ability of CTCF and Sp1 transcription factors from binding to their target sites within the promoter suggests that alternative mechanisms are possible (Ben-Hattar *et al.*, 1989). Furthermore, it can take 48 hours to repress methylated reporter constructs suggesting that inhibition of transcription might be mediated by indirect mechanisms that take time to be established in the cell (Buschhausen *et al.*, 1985). In line with this, the establishment of chromatin is required for DNA methylation-dependent inhibition (Buschhausen *et al.*, 1987), this chromatin may spread along a plasmid to cover unmethylated sites (Kass *et al.*, 1993), and repression of transcription can occur when the DNA methylation is placed distally from the gene of interest (Nan *et al.*, 1997). This data suggests that repression can be achieved after MBD-containing proteins have recognised and bound to the mCpG mark in order to establish a

repressive environment. These proteins may then function in a variety of manners; inhibition of the binding of transcription factors, interference with the binding and/or progression of RNA polymerase or alterations of the chromatin architecture to create a general inactive chromatin environment by recruitment of histone modifying enzymes. In line with this, histone modifications that are generally associated with gene repression are often found co-localised with sites of methylated CpGs, such as H3K9me3, whereas unmethylated CGIs are associated with active histone modification marks such as H3K4me3 (Illingworth *et al.*, 2010; Maunakea *et al.*, 2010; Weber *et al.*, 2007).

While DNA methylation and gene repression are certainly linked, most studies described so far demonstrate this relationship by looking at promoters, which *in vivo* are predominantly unmethylated. How the bulk of the vertebrate genome methylation (which is situated in gene bodies and intergenic regions) affects transcription is unknown. Gene-body methylation of an integrated reporter construct in mouse cells has been shown to reduce transcription compared to an unmethylated construct, despite the promoter remaining methylation-free, thus demonstrating that gene-body methylation has an ability to repress transcription, independent of the promoter (Appanah *et al.*, 2005). A hypothesis for the function of endogenous gene-body methylation is that it works to repress transcription in such a manner as to reduce transcriptional noise (Bird *et al.*, 1995). This hypothesis stems from the fact that invertebrate genomes lack DNA methylation, whereas vertebrate genomes maintain very high levels of it, despite its propensity to induce sequence mutations (Tweedie *et al.*, 1997). Vertebrate genomes have larger numbers of genes and highly complex developmental patterns which perhaps require an extra constraint on transcription to prevent inappropriate expression at the wrong time or place during development. Evidence to support such a mechanism includes the fact that intragenic DNA methylation has a role in the regulation of alternative promoters in different cell types (Maunakea *et al.*, 2010), and that upon depletion of DNA methylation, transcription does not dramatically increase, as might be expected, but minor changes occur (Feng *et al.*, 2010). A mechanism for how gene-body methylation might interfere with spurious transcription and enforce a tight control on inappropriate transcription is unknown. While the results of Appanah *et al* indicated that RNAP II recruitment to the promoter of a gene-body methylated transgene in mouse cells was reduced compared to an unmethylated transgene (Appanah *et al.*, 2005), nuclear run-on experiments in the fungus *Neurospora crassa* found that methylation reduces the rate of transcription by 5-fold by inhibiting RNAP elongation rather than RNAP initiation (Rountree & Selker, 1997). Thus recent observations that high levels of gene body methylation correlates with

increased gene expression could be explained by gene body methylation functioning to prevent inappropriate transcription throughout the gene by interfering with RNAP initiation and/or elongation (Ball *et al.*, 2009; Flanagan & Wild, 2007; Rauch *et al.*, 2009). Perhaps the function of DNA methylation, and of the proteins that bind to this modification, is to dampen gene expression and provide an extra constraint on inappropriate RNAP function, thus providing an environment where expression of appropriate full-length transcripts can be achieved.

Thus the question still remains, how do proteins that bind to mCpG bring about gene repression or transcriptional restraint in a cell- and locus-specific manner? Is the action of occupying sites of methylation enough to inhibit RNA polymerase initiation and/or elongation, or are downstream events also required. If so, what are these downstream events?

1.1.4 DNA methylation and the mammalian brain

DNA methylation is known to be dynamically regulated during mammalian development, and in particular in the brain DNA methylation dynamics are associated with neuronal function, such as long-term potentiation (LTP) [for reviews see: (Graff *et al.*, 2011; Telese *et al.*, 2013)]. In line with the importance of DNA methylation for proper neuronal function, post-natal deletion of DNMT1 and DNMT3A in forebrain excitatory neurons results in learning, memory, and synaptic plasticity defects (Feng *et al.*, 2010), while CNS deletion of DNMT3A results in sever motor defects and mice die prematurely (Nguyen *et al.*, 2007).

While methylation of the CpG dinucleotide has been the predominant focus of the field, recent analysis of non CpG methylation (i.e. mCpH) has discovered widespread deposition of this mark *in vivo,* with evidence for a role in gene regulation. Initial *in vitro* studies found that all three DNMT enzymes can methylate CpH sites (Hubrich-Kuhner *et al.*, 1989; Yokochi & Robertson, 2002), which led to subsequent analysis of mCpH *in vivo*. This mark, in particular mCpA, is found in the male and female mouse germ line (Ichiyanagi *et al.*, 2013; Smith *et al.*, 2012; Tomizawa *et al.*, 2011), mouse and human ESCs (Laurent *et al.*, 2010; Ramsahoye *et al.*, 2000; Ziller *et al.*, 2011) and human iPSCs (Ziller *et al.*, 2011), however was not believed to be present in differentiated cell types due to its absence from human fibroblasts, embryoid bodies, neural precursor cells, pancreatic islets, rectal mucosa, rectal smooth muscle, skeletal muscle, stomach muscle and blood samples (Laurent *et al.*, 2010; Ziller *et al.*, 2011), as well as being absent from mouse liver, kidney, spleen and lung (Ramsahoye *et al.*, 2000). Recent analysis of mouse and human brain samples however have

revealed that a significant proportion of cytosine methylation in the adult brain is composed of the CpH dinucleotide (Guo *et al.*, 2013; Lister *et al.*, 2013; Varley *et al.*, 2013; Xie *et al.*, 2012). As a proportion of all cytosine methylation, mCpH makes up 25% in the dentate gyrus (Guo *et al.*, 2013), and 53% in neurons of the frontal cortex (Lister *et al.*, 2013). The levels of CpH methylation increase during development, reaching a peak of 1.5% and 1.3% mCpH/CpH in human and mouse adults respectively (Lister *et al.*, 2013). In particular the most rapid increase is during the first two years of post-natal human development, a time period that coincides with rapid levels of synaptogenesis and increases in synaptic density (Lister *et al.*, 2013). These high levels of non-CpG methylation in the adult central nervous system, and the absence of it from most other differentiated cell types suggests a unique role for CpH methylation in the function of neurons.

The function of CpH methylation in neurons appears to be linked to gene repression. Gene expression is inversely correlated with mCpH in the gene body and the surrounding ±5 kb environment and in particular, lowly expressed genes tend to have higher gene body mCpH levels than their surrounding regions suggesting a role for gene body methylation in gene repression (Guo *et al.*, 2013). Furthermore, genes that lose mCpG and mCpH during neuronal development tend to be highly expressed, constitutive genes that are enriched for neuronal functions and depleted of astrocytic functions. These genes concomitantly gain H3K4me1 and H3K27ac marks as well as exhibit increased DNaseI hypersensitivity, whereas genes that gain mCpH during development tend to have lower levels of H3K4me1, H3K27ac and DNaseI hypersensitivity (Lister *et al.*, 2013). This indicates that the loss of CpH methylation during development occurs concomitantly with the gain of active chromatin marks indicative of active transcription. While there is a tendency for binding sites of neuron-specific transcription factors to be depleted of mCpH (Guo *et al.*, 2013), it remains to be determined whether active transcription prevents mCpH deposition, or encourages active mCpH removal. Nevertheless, transfection of mCpH-methylated reporter gene constructs into mouse hippocampal neurons is sufficient to significantly repress transcription to similar levels as mCpG does (Guo *et al.*, 2013).

Efforts have also been made to establish the proteins that deposit or bind to this modification. While DNMT3B has been suggested to be responsible for mCpH in ESCs due to its high expression levels (Laurent *et al.*, 2010), DNMT3A is the more likely candidate for mCpH deposition in neurons. The first evidence came from bioinformatics analysis that showed a tendency for DNMT3A sites in the neuronal genome to have higher levels of CpH

methylation compared to a random genomic sampling (Guo *et al.*, 2013; Lister *et al.*, 2013). Using adeno-associated viruses (AAVs) to deliver shRNAs against DNMT3A into the adult mouse dentate gyrus demonstrated a significant decrease in the level of CpH methylation, with no effect on mCpG levels, and this resulted in a loss of repression at the loci that lost gene body mCpH, providing strong evidence for the role of DNMT3A as the depositor of mCpH in neurons, and for the role of mCpH in transcriptional repression (Guo *et al.*, 2013). This study also identified MeCP2 as a specific binder of mCpH. *In vitro* EMSA assays found that MeCP2, but not MBD2, could bind to mCpH, albeit at a slightly reduced capacity than for mCpG. This was supported with *in vivo* ChIP analysis which showed that MeCP2-immunoprecipatated DNA fragments contained mCpH as determined by bisulphite analysis (Guo *et al.*, 2013). As MeCP2 is the major focus of this study, further information about the binding specificities of MeCP2 will be described in Chapter 1.2.4.

The discovery that mCpH is highly enriched in the neuronal genome and the fact that the levels of this mark are less binomial than mCpG strongly suggests a role for mCpH in the regulation of transcriptional networks during mammalian brain development. Furthermore, the strong conservation of methylation patterns between human and mouse suggests a functional role (Maunakea *et al.*, 2010). In line with this, post-natal deletion of all three DNMT enzymes in the postnatal hippocampus led to a decrease in the level of mCpH with little effect on mCpG, suggesting that mCpH has the potential for dynamic regulation even within a differentiated cell type (Guo *et al.*, 2013).

1.2 Methyl-CpG binding protein 2

1.2.1 A brief introduction to MeCP2

MeCP2 (methyl-CpG binding protein 2) was discovered in 1992 by Southwestern assays using rat brain extracts that were designed to identify methyl-CpG binding proteins (MeCPs) that could specifically bind to methylated DNA (Lewis *et al.*, 1992). It was identified with an apparent molecular weight of 84 kDa (Lewis *et al.*, 1992), showed highest expression (as judged by Southwestern assays) in rat and mouse brain tissue compared to other tissues (Meehan *et al.*, 1992), and is tightly bound to chromatin, requiring high salt to remove it from DNA (Meehan *et al.*, 1992). Deletion analysis of full-length rat MeCP2 protein identified the minimal region required for binding to methylated DNA, termed the methylated DNA binding domain (MBD) which consists of amino acids 78-162 and has a footprint of 12-14 bp around a single mCpG site (Nan *et al.*, 1993). Further deletion analysis

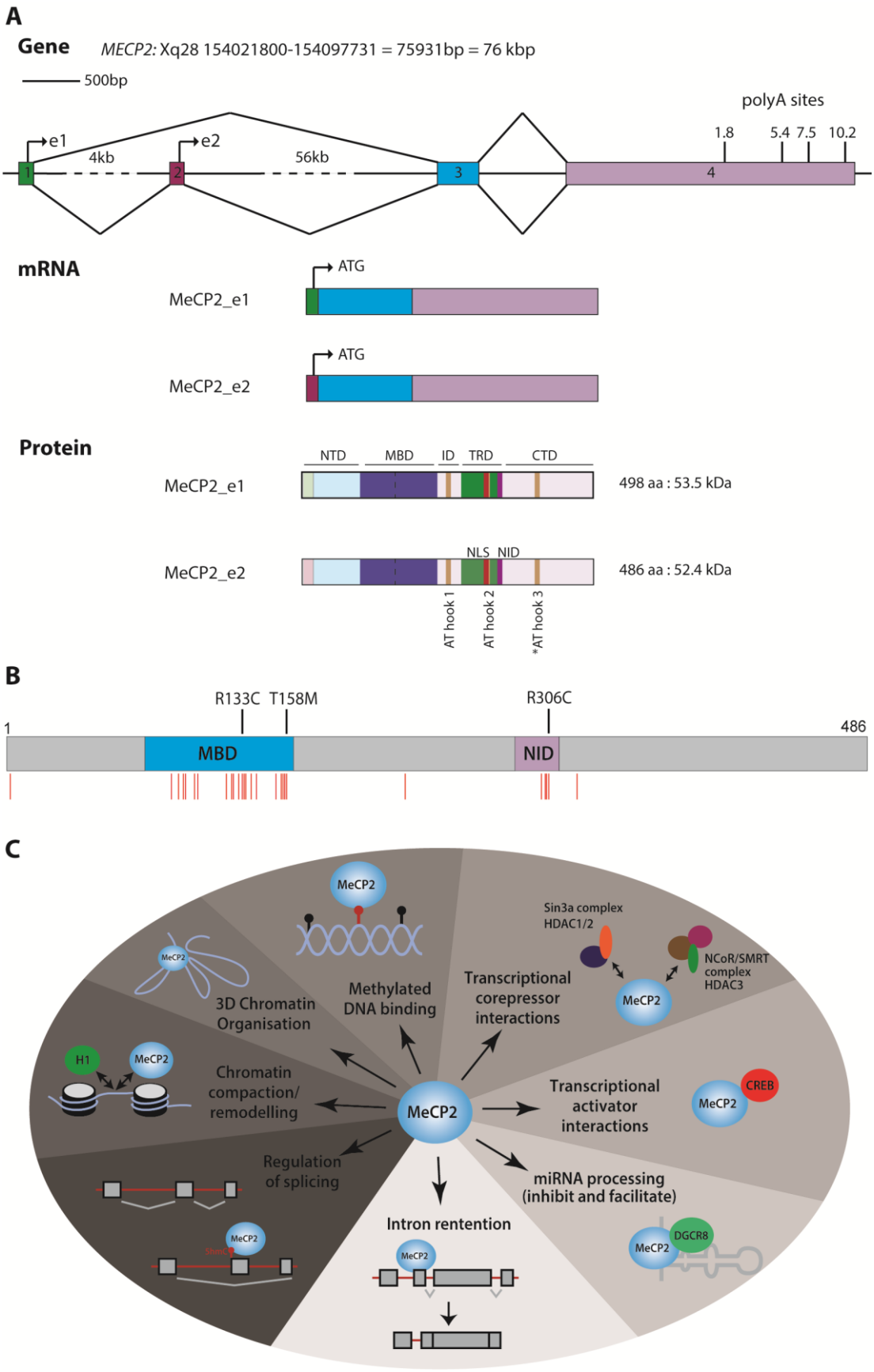
identified a nuclear localisation signal (NLS) that corresponds to amino acids 255-271 (Nan *et al.*, 1996). The observation that MeCP2 binds specifically to methylated DNA suggested that it may play a role transcriptional repression. In line with this, a transcription repression domain (TRD), defined as the minimal region required to repress transcription from a reporter gene when transfected into cells, was discovered that consists of amino acids 207-310 (Nan *et al.*, 1997) A small region within the TRD was later identified to be responsible for an interaction with the NCoR/SMRT corepressor complex, and was termed the NCoR Interaction Domain (NID) (Lyst *et al.*, 2013). The importance of this small domain and the MBD are exemplified by the fact that missense mutations within these domains cause a neurological disorder called Rett syndrome (RTT), whereas most missense mutations outwith these two domains are tolerated in the general population as judged by the ExAC database (Figure 1B) (Amir *et al.*, 1999; Lyst *et al.*, 2013). This has led to the "bridge" hypothesis whereby the function of MeCP2 is to recruit the NCoR/SMRT complex to sites of methylated DNA in order to maintain gene repression. Evidence for and against this hypothesis will be discussed in the rest of Chapter 1.2.

1.2.2 Gene and protein structure

When MeCP2 was first purified from rat brains in 1992, digestion followed by partial sequencing of the amino acid sequence allowed for primers to be designed that ultimately led to the sequencing and cloning of full-length, rat *Mecp2* cDNA (Lewis *et al.*, 1992). Two years later the mouse *Mecp2* gene was mapped to the X-chromosome (Quaderi *et al.*, 1994), and two years after that the precise location was determined in human cells and was shown to be subject to X inactivation, suggesting altered dynamics of regulation between males and

Figure 1. Gene, mRNA and protein schematics of *MECP2*/MeCP2.

(A) Schematics of the *MECP2* gene, *MECP2* mRNA and MeCP2 protein structures. The *MECP2* gene is composed of 4 exons and has 4 alternative polyadenylation sites which produce transcripts of different sizes. Gene location coordinates are from GRCh38.p7 assembly. Two mature *MECP2* mRNAs are produced, one being encoded from exon 1 and the other from exon 2. Two different MeCP2 protein isoforms are expressed. The shaded colours in the protein schematic are reflective of the portion of the gene that encodes that portion of protein. The MeCP2 protein can be roughly divided into the N-terminal domain (NTD), the methylated DNA binding domain (MBD), the intervening domain (ID), the transcriptional repression domain (TRD) and the C-terminal domain (CTD). The MBD is shown in dark blue, the TRD is shown in dark green, the NCoR/SMRT interaction domain (NID) is shown in dark purple, the nuclear localisation signal (NLS) is shown in red and three putative AT-hooks are shown in orange. Note that AT-hook 2 overlaps with the NLS. The asterisk on AT-hook 3 indicates that this sequence is not the consensus for a true AT-hook. **(B)** Protein schematic of MeCP2-e2 showing the MBD and NID and the location of RTT-causing missense mutations in red, with the 4 missense mutation that are studied in this project labelled. Schematic altered from (Shah & Bird, 2017). **(C)** Schematic illustrating the plethora of protein interactions and cellular functions that have been proposed for MeCP2.


A
Gene
MECP2: Xq28 154021800-154097731 = 75931bp = 76 kbp
500bp
polyA sites
e1
e2
4kb
56kb
1.8
5.4 7.5
10.2
1
2
3
4
mRNA
ATG
MeCP2_e1
ATG
MeCP2_e2
Protein
NTD
MBD
ID
TRD
CTD
MeCP2_e1
498 aa : 53.5 kDa
NLS NID
MeCP2_e2
486 aa : 52.4 kDa
AT hook 1
AT hook 2
*AT hook 3
B
R133C T158M
R306C
1
486
MBD
NID
C
MeCP2
3D Chromatin Organisation
Methylated DNA binding
Sin3a complex HDAC1/2
NCoR/SMRT complex HDAC3
Transcriptional corepressor interactions
H1
Chromatin compaction/ remodelling
Transcriptional activator interactions
CREB
Regulation of splicing
miRNA processing (inhibit and facilitate)
Intron rentention
DGCR8

females and identifying a possible role for MeCP2 in disorders involving the X-chromosome (D'Esposito *et al.*, 1996). Detailed sequence analysis of both the human and mouse transcripts revealed four alternative polyadenylation (polyA) signals within a large 3' untranslated region (UTR), suggesting the potential for tissue- and developmental-specific regulation of MeCP2 expression levels (Figure 1A) (Coy *et al.*, 1999; Reichwald *et al.*, 2000). In line with this, recent studies have found that the levels of the two main alternative transcripts (the 10.2 kb and 1.8 kb transcripts) are tightly regulated during development as well as between the central nervous system and peripheral tissues, leading to high expression levels in post-mitotic neurons that increases during development (McGowan & Pang, 2015; Pelka *et al.*, 2005; Rodrigues *et al.*, 2016).

For many years *MECP2* was thought to contain only three exons, all of which contributed to the coding sequence. In 2000, sequence analysis of the human and mouse loci revealed a novel upstream exon (Reichwald *et al.*, 2000), which was later shown to be a coding exon, thus identifying a novel protein isoform of MeCP2 which has an alternative N-terminus (Kriaucionis & Bird, 2004; Mnatzakanian *et al.*, 2004). The MeCP2_e1 isoform (previously MeCP2B or MeCP2α) begins transcription in exon 1, creating a unique 32 amino acid N-terminus, and splices out exon 2 from the mature mRNA transcript, while the MeCP2_e2 isoform (previously MeCP2A or MeCP2β) is encoded from exon 2 creating a shorter, 20 amino acid N-terminus (Figure 1) (Kriaucionis & Bird, 2004). Both isoforms contain the coding sequence from exons 3 and 4 which contain the MBD and TRD, and thus it is not surprising that both isoforms localise to methylation-dense heterochromatic foci in mouse cells (Kriaucionis & Bird, 2004). The fact that human MeCP2_e1 alone is able to rescue phenotypes in *Xenopus laevis* embryo's (whose endogenous MeCP2 is more similar to MeCP2_e2) suggests that both isoforms have the same function (Stancheva *et al.*, 2003). Furthermore, the phenotype in *Mecp2*$^{-/y}$ mice can be rescued by either isoform (Kerr *et al.*, 2012). Yet the observation that a RTT-causing mutant (A2V) is found in exon 1 of *MECP2* and is therefore only specific to the MeCP2_e1 isoform (Fichou *et al.*, 2009), suggests differential importance of the two isoforms. Indeed *in vivo* studies have found the MeCP2_e1 isoform to be the more highly expressed isoform in a range of mouse tissues, in the human brain, and during mouse development (Dragich *et al.*, 2007; Kriaucionis & Bird, 2004) thus explaining the debilitating effect of the e1-specific A2V RTT mutation. While MeCP2_e1 is now known to be the major isoform expressed *in vivo*, all RTT mutation nomenclatures were designated according to the MeCP2_e2 isoform. Thus, for example, the most common RTT mutation T158M corresponds to threonine 158 in MeCP2_e2 which is

threonine 170 in the expressed MeCP2_e1 isoform. Throughout this thesis, all RTT mutations will be referred to using the standard MeCP2_e2 nomenclature.

When MeCP2 protein was first identified, analysis of the 492 amino acid (aa) long rat protein revealed some interesting features that have implications for the DNA binding properties of MeCP2 (Lewis *et al.*, 1992). Firstly, MeCP2 is a very basic protein with lysine and arginine residues comprising 22.5% of the entire MeCP2 protein sequence. Secondly, MeCP2 contains two (R)GRP(K) motifs which in other proteins are known to be responsible for binding to AT-rich DNA sequences (AT hooks, see Figure 1)), and eight SPKK motifs that are implicated in binding to the minor groove of AT-rich DNA. Finally, the methylated-DNA binding domain also contributes to DNA binding specificities (Lewis *et al.*, 1992; Nan *et al.*, 1993). A highly basic protein that contains multiple motifs for binding to AT-rich or methylated DNA suggests that binding to DNA is a major function of MeCP2.

Efforts towards solving the structural conformation of MeCP2 protein have been hampered by some unusual protein characteristics. Despite the amino acid sequence predicting a protein of roughly 53 kDa in size, MeCP2 migrates in an SDS-PAGE gel at ~75-80 kDa, and size exclusion chromatography finds MeCP2 eluting at 400-500 kDa (Klose & Bird, 2004). This unexpected pattern is not due to multimers because band shifts using tagged and untagged MeCP2 and analytical centrifugation suggests that MeCP2 is a monomer (Adams *et al.*, 2007; Nan *et al.*, 1993). The precise reason for the unexpected migration pattern was assumed to be caused by the rigid, rod-like structure (Adams *et al.*, 2007) and the elongated shape that MeCP2 exhibits (Klose & Bird, 2004), and this was later confirmed by small angle X-ray scattering (Yang *et al.*, 2011).

Circular dichroism (CD) spectroscopy has demonstrated that approximately 59% of the protein is unstructured (40% of the MBD and 85% of the TRD are also unstructured), highlighting that in-depth structural analysis of full-length MeCP2 will be a challenging process (Adams *et al.*, 2007). Trypsin digestion studies to identify "weak points" in the MeCP2 structure revealed that the MBD, the TRD and the C-terminal domain of MeCP2 do retain some tertiary structure that protects them from proteolytic cleavage (Adams *et al.*, 2007), and in line with this the MBD and the NID are the only two structures that have been solved to date. The first structure was the solution structure of the MBD (amino acids 77-164) which was solved using NMR spectroscopy and revealed for the first time, the DNA-binding interface of the MBD of MeCP2 (Wakefield *et al.*, 1999). The MBD tertiary

structure was described as a non-symmetrical, wedge-shape that positions multiple lysine and arginine residues at the surface thus creating a large non-specific DNA recognition interaction surface. Later came the crystal structure of the MBD (amino acids 77-167) complexed to a 20bp DNA probe (Ho *et al.*, 2008). This highlighted the importance of water molecules for recognition of the mCpG dinucleotide and uncovered an Asx-ST turn that is critical for structural integrity of the MBD. Interestingly, individual residues that were highlighted as being important for mCpG recognition and/or MBD structure were found to be frequently mutated in RTT (Ho *et al.*, 2008). More recently a co-crystal structure of a short NID peptide was solved in complex with the C-terminus of the TBL1R1 protein, an NCoR complex subunit (Kruusvee *et al.*, 2017). Importantly, the four individual residues in the NID that make contact with TBL1R1 are all mutated in RTT cases and, looking at the interface from the reverse angle, two important residues in TBL1R1 for MeCP2 recognition are frequently found mutated in developmental delay and intellectual disorders (Kruusvee *et al.*, 2017). Interestingly, a TBL1R1 residue (Y446) that when mutated causes the intellectual disability disorder Pierpoint syndrome (Y446C) (Heinen *et al.*, 2016), was found at the NID-interaction interface. Mutation of Y446 to phenylalanine, however, did not disrupt the interaction with MeCP2 (Kruusvee *et al.*, 2017). Perhaps analysis of the exact Pierpoint Y446C mutation might reveal a deficit in the interaction with MeCP2 or alternatively Pierpoint syndrome might be caused by the disruption of TBL1R1 with other interaction partners.

1.2.3 Expression pattern of MeCP2 during mammalian development

When MeCP2 was discovered, analysis of protein expression in different mouse tissues was assessed by Southwestern assay which found it to be highly expressed in mouse and rat brain, with lower levels in mouse spleen, liver and kidney, and the lowest levels in rat testes (Meehan *et al.*, 1992). Assessment of transcript levels found *MECP2* ubiquitously in a range of human tissues (Coy *et al.*, 1999; D'Esposito *et al.*, 1996; Reichwald *et al.*, 2000), but analysis of protein levels by immunofluorescence and quantitative laser scanning cytometry determined protein expression to be highest in central nervous system (CNS) tissues in both mouse and human post-mortem samples (Balmer *et al.*, 2003; LaSalle *et al.*, 2001). A broad distribution of MeCP2 expression was observed in the CNS tissues, and when cells were distinguished according to neuron vs glial status, glial cells uniformly contained lower amounts of MeCP2, while neuronal cells still displayed a broad distribution of MeCP2 abundance (Balmer *et al.*, 2003; LaSalle *et al.*, 2001). This result was confirmed by Western blot analysis of mouse tissues and immunofluorescence analysis of mouse and human brain slices, which also confirmed the low-to-absent expression of MeCP2 in astroglia and

oligodendrocytes (Kishi & Macklis, 2004; Shahbazian *et al.*, 2002a). More recently, semi-quantitative analysis of MeCP2 levels in adult mouse tissues has further established that MeCP2 expression is highest in the brain (although within the brain, expression is low within the cerebellum), and is much lower in other tissues (Ross *et al.*, 2016).

Analysis of MeCP2 expression during mouse development demonstrated an increase in levels and strong punctate staining, corresponding to the localisation of MeCP2 at heterochromatic foci in mouse nuclei, which is where the majority of mC in the mouse genome is localised (Kishi & Macklis, 2004; Shahbazian *et al.*, 2002a). Immunohistochemistry in rat brain slices also found high levels of expression in neurons, an increase of expression during development, and low levels in astrocytes (Jung *et al.*, 2003). These results have also been confirmed by tissue culture neuronal differentiation protocols using rat, mouse and human cells (Jung *et al.*, 2003; Kishi & Macklis, 2004). More recently, quantitative approaches have determined the number of MeCP2 molecules in the adult mouse NeuN+ nucleus to be 16 x 10^6, while NeuN- cells only have approximately 2 x 10^6 molecules/nucleus. This high level of expression in mature neurons of the CNS puts MeCP2 expression at near histone octamer levels (Skene *et al.*, 2010).

Thus, MeCP2 is ubiquitously expressed, particularly at the transcript level, but post-transcriptional regulatory mechanisms act to increase the expression level significantly in neurons of the central nervous system. The expression level of MeCP2 in neurons increases as development progresses, and MeCP2 staining becomes more punctate, indicative of strong hetereochromatic binding.

1.2.4 MeCP2 as a DNA-binding protein

There is extensive evidence for the ability of MeCP2 to bind to methylated DNA. These include Southwestern and band shift assays that established that symmetrically modified mCpG is a binder of MeCP2, hemi-methylated DNA cannot bind MeCP2 and binding to non-CpG methylation (mCpH) is observed but is extremely weak (Jones *et al.*, 1998; Lewis *et al.*, 1992; Meehan *et al.*, 1992). Furthermore methylation of the guanine residue in the mCpG dinucleotide has no effect on MeCP2 binding (Nan *et al.*, 1993).

This *in vitro* data is supported by immunofluorescence analysis, where in mouse cells MeCP2 co-localises with heterochromatic foci in interphase cells and with pericentromeric repeats in metaphase spreads, both of which contain high mCpG densities in the genome

(Lewis *et al.*, 1992). Furthermore, MeCP2 localisation to heterochromatic foci is lost in mouse cells that have ~5% the level of m^5C due to a disrupted DNMT1 gene (Nan *et al.*, 1996). Immunofluorescence analysis of MeCP2 localisation in other organisms suggested early on that MeCP2 might bind globally throughout the genome as human HeLa, monkey COS7 and rat cells all show broad distribution throughout chromosome arms which, in comparison to mouse cells, do not have mCpG-rich pericentromeric heterochromatin (Nan *et al.*, 1997).

ChIP experiments have further confirmed the mCpG-binding preferences of MeCP2 *in vivo*. MeCP2 associates with integrated proviral DNA in a methylation-specific manner in mouse cells (Lorincz *et al.*, 2001), binds to a hypermethylated viral promoter in a 5-aza-cytidine dependent manner in chicken cells (Rietveld *et al.*, 2002), and 5-aza-cytidine treatment in neural stem cells dramatically reduces the association of MeCP2 with a methylated reporter construct (Muotri *et al.*, 2010). While these experiments use reporter transgenes and integrated viral DNA, there is also evidence at endogenous loci that MeCP2 associates with DNA in a methylation-dependent manner. For example ChIP in mouse neurons reveals that in female samples, MeCP2 specifically associates with the *XIST* allele that is methylated, rather than the unmethylated allele, (Skene *et al.*, 2010), and the endogenous *MDR1* CGI in human T cells loses MeCP2 association upon 5-aza-cytidine treatment (El-Osta *et al.*, 2002). Similarly, MeCP2 ChIP-on-ChIP signal is lost when human MRC5 cells are treated with 5-aza-cytidine (Klose *et al.*, 2005).

Genome-wide ChIP-sequencing studies have recently assessed the profile of MeCP2 binding throughout the genome. As might be expected from a protein whose binding site is abundant in the genome, and based on initial immunofluorescence analysis (Nan *et al.*, 1997), MeCP2 binding was found throughout the entire genome, with the most prominent feature being a lack of binding at unmethylated CGIs in mouse whole brain (Skene *et al.*, 2010), cortex and cerebellum (Gabel *et al.*, 2015), hypothalamus (Chen *et al.*, 2015) and cultured mouse neurons (Cohen *et al.*, 2011). Surprisingly, such genome-wide analysis uncovered an enrichment of MeCP2 binding in areas of low mCpG density. These areas, however, have a high mCpH density and further analysis found that genes containing high levels of gene-bosy mCpH have high levels of MeCP2 occupancy (Chen *et al.*, 2015). As detailed in Chapter 1.1.4, mCpH levels increase during mammalian brain development and this increase coincides with the increase in MeCP2 expression during development (Chapter 1.2.3). The mCpH dinucleotide was refined to a mCpA dinucleotide by EMSA and ChIP-seq analysis

(Gabel *et al.*, 2015), agreeing with earlier band shift experiments that also identified non-mCpG binding capabilities of MeCP2 (albeit with a very weak affinity in the earlier study) (Guo *et al.*, 2013; Meehan *et al.*, 1992). The most recent analysis of MeCP2 binding by EMSA and ChIP-seq analysis has refined the mCpH preference to mCAC and modelled the potential binding interface of the MBD with a mCAC motif by *in silico* analysis of the crystal structure of the MBD (Lagger *et al.*, 2017).

Attempts to identify binding preferences outwith the methylated residue include methyl-SELEX studies which found a preference for an [A/T] run of 4 or more bases adjacent to the mCpG dinucleotide (Klose *et al.*, 2005), a result that was strengthened by the X-ray crystal structure of the MBD which showed a run of [A/T] residues induced a DNA bend that could accommodate DNA binding by the Asx-ST motif in MeCP2 (Ho *et al.*, 2008).

The evidence finds that MeCP2 binds to mCpG and mCAC *in vitro* and *in vivo,* and demonstrates how genes that have high levels of these two modifications (and therefore high levels of MeCP2 occupancy) in WT cells are the most misregulated in MeCP2 KO cells (Chen *et al.*, 2015; Gabel *et al.*, 2015; Lagger *et al.*, 2017). The extremely high incidence of these two modifications throughout the genome (~1 every 100bp) has implications for the mechanism of gene repression by MeCP2. It implies that MeCP2 will affect the expression of thousands of genes to act as a global repressor. Quite how MeCP2 might bring about global transcriptional repression or exert a restraint on gene expression is unknown. Even though MeCP2 has the capacity to bind to a large number of genes, likely the local transcriptional and chromatin environment of each individual gene will impact upon the ability of MeCP2 to bind to and regulate that gene.

1.2.5 MeCP2 as a transcriptional repressor

When MeCP2 was first identified a methylation-specific gene repression role could not be found. *In vitro* experiments using purified rat MeCP2, the human α-globin locus and rat liver nuclear extracts failed to determine any mCpG-dependent transcriptional repression. In fact, transcription was inhibited regardless of the methylation status of the α-globin locus. Thus, it was assumed that MeCP2 regulates gene expression by simply binding to DNA in a non-discriminate manner (Meehan *et al.*, 1992). A later study using a purer MeCP2 protein preparation managed to demonstrate methylation-specific repression activities of MeCP2 *in vitro* (Nan *et al.*, 1997). This study also mapped the minimal region required to mediate repression when tethered to loci using GAL4 domains (amino acids 207-310, see Figure 1),

and demonstrated that an N-terminal peptide of amino acids 1-162 could weakly repress transcription *in vitro*, suggesting that binding of MeCP2 to promoters was not enough for full repression and that the activity of the so-called transcriptional repression domain (TRD) was required for complete repression. Methylation-dependent repression by MeCP2 *in vitro* was confirmed using MeCP2 purified from *Xenopus* (Kaludov & P.Wolffe, 2000), and both studies found that MeCP2 could repress transcription from a distance, either 1.3kb and 2.1kb upstream of the promoter (Nan *et al.*, 1997) or 250bp and 500bp upstream (Kaludov & P.Wolffe, 2000). These experiments are also supported by luciferase reporter assays in mouse tail fibroblasts that demonstrate MeCP2 can preferentially repress a methylated luciferase gene (Guy *et al.*, 2001), and in human HeLa cells where MeCP2 can repress transcription in a methylation-dependent manner and this repression is abolished upon deletion of the TRD (but surprisingly not with the R306C mutation, see discussion below) (Drewell *et al.*, 2002).

The observation that the binding of a TRD-less MeCP2 could not fully repress a reporter construct (Nan *et al.*, 1997) led to the hypothesis that the function of the TRD was to recruit a corepressor complex to DNA to mediate full repression. Many such protein interactions have been identified over the years including Sin3a (Nan *et al.*, 1998), Ski (Kokura *et al.*, 2001), NCoR (Kokura *et al.*, 2001), SMRT (Stancheva *et al.*, 2003), CoREST (Lunyak *et al.*, 2002), SUV39H1 (Lunyak *et al.*, 2002), YY1 (Forlani *et al.*, 2010) and TBLR1 (Kruusvee *et al.*, 2017). The majority of these interactions have not been assessed in multiple organisms or by different laboratories and the *in vivo* relevance of many of these findings remains to be determined (Lyst & Bird, 2015). For many years Sin3a was to be considered the major interactor of MeCP2 and interactions were demonstrated in HeLa cells (Nan *et al.*, 1998), rat brain extracts (Nan *et al.*, 1998) and *Xenopus* oocyte extracts (Jones *et al.*, 1998). The observation that MeCP2-mediated repression was reduced upon treatment with an HDAC inhibitor (TSA) was consistent with Sin3a being the transcriptional mediator as Sin3a is known to interact with HDAC1 and HDAC2 (Jones *et al.*, 1998; Nan *et al.*, 1998). The importance of Sin3a interactions for *in vivo* function, however, was questioned as the interaction domain was found to not overlap with the TRD (Nan *et al.*, 1998), and more importantly the interaction was found to be rather unstable (Klose & Bird, 2004) and in fact quite weak (Lyst *et al.*, 2013).

Since then the focus has shifted to the NCoR/SMRT corepressor complex which was first identified as a potential MeCP2 interacting complex *in vitro* in 2001 (Kokura *et al.*, 2001)

and 2003 (Stancheva *et al.*, 2003), but the importance of this interaction was not confirmed until 2013 (Lyst *et al.*, 2013). The NCoR/SMRT corepressor complex is a large, multi-subunit complex predicted to contain around 10-12 different protein components, including the histone deacetylase HDAC3. It functions to repress transcription through its association with nuclear receptors such as thyroid hormone receptors and retinoic acid receptors, with the repression being dependent on NCoR, SMRT, HDAC3, TBL and TBLR1 activities (although due to significant homology between TBL1 and TBLR1 it is perhaps not surprising that repression is alleviated only when both components are removed from the system, suggesting redundancy) (Yoon *et al.*, 2003a). Homozygous deletion of NCoR from mice found the protein to be necessary for embryonic development (with embryos dying by E15.5) while T cell development, erythropoiesis and CNS development (specifically neural precursor self-renewal) were all impaired (Hermanson *et al.*, 2002; Jepsen *et al.*, 2000). This study also found the NCoR complex to have varying functions in repression and activation depending on the co-receptors, ligands and loci in question (Jepsen *et al.*, 2000). Of note, NCoR has been implicated in binding to another mCpG-binding protein Kaiso both *in vitro* and *in vivo* and is required for Kaiso-mediated repression of the *MTA2* locus *in vivo* in HeLa cells (Yoon *et al.*, 2003b). Thus the NCoR/SMRT corepressor complex has been identified in a number of *in vivo* complex interactions suggesting it may play roles in the regulation of a number of different loci under a variety of different conditions.

The interaction of MeCP2 with the NCoR/SMRT complex has been mapped to a C-terminal portion of the MeCP2 TRD domain, termed the NID (NCoR-interaction domain, Figure 1) (Lyst *et al.*, 2013). This interaction is abolished when a NID-containing RTT mutation R306C is introduced and furthermore mice containing this mutation exhibit significant RTT-like phenotypes, similar to MeCP2 null mice, and have a decreased lifespan (Lyst *et al.*, 2013). Furthermore, all four RTT mutations that localise to the NID domain are able to abolish the interaction of MeCP2 with NCoR and result in reduced repression of a luciferase construct in mouse NIH-3T3 assays highlighting the functional significance of this MeCP2-NCoR interaction for MeCP2 function *in vivo* (Lyst *et al.*, 2013). The specific MeCP2 interaction site on the NCoR/SMRT complex has recently been mapped to the WD40 domain of the TBL1/TBLR1 proteins with a co-crystal structure identifying the direct interactions the NID residues have with TBLR1. Interestingly, TBLR1/TBL1 mutations that cause intellectual disability and autism spectrum disorders map to the MeCP2-TBL interface and disrupt MeCP2 binding (Kruusvee *et al.*, 2017).

While for many years repression assays confirmed the hypothesis that MeCP2 functions as a transcriptional repressor, global transcriptome analysis did not find an overall upregulation of gene expression upon MeCP2 deletion. There have been numerous microarray studies of MeCP2 null mouse brain regions at different ages of phenotype progression, all of which found small numbers of significantly changing genes that both increase and decrease in their expression level (Ben-Shachar *et al.*, 2009; Chahrour *et al.*, 2008; Jordan *et al.*, 2007; Kriaucionis *et al.*, 2006; Nuber *et al.*, 2005; Tudor *et al.*, 2002; Urdinguio *et al.*, 2008). The largest changes were observed in the study by Chahrour and colleagues who looked at the hypothalamus and found 2582 genes that were misexpressed in both MeCP2 null and transgenic (overexpression) animals (Chahrour *et al.*, 2008). Indeed, microarray analysis of patient samples have revealed the same patterns of low numbers of genes that increase or decrease their expression level less than 2-fold in the cerebral cortex, superior frontal cortex, T lymphocytes, B lymphoblastoid cells and fibroblasts (Ballestar *et al.*, 2005; Colantuoni *et al.*, 2001; Delgado *et al.*, 2006; Deng *et al.*, 2007; Nectoux *et al.*, 2010; Traynor *et al.*, 2002). As well as low numbers of changing genes, there is no clear functional group of genes that are misregulated in every sample and each study seemed to find its own group of "MeCP2 target genes", for example glucocorticoid-related genes (Nuber *et al.*, 2005), G-protein coupled receptors (Chahrour *et al.*, 2008), mitochondrial genes (Kriaucionis *et al.*, 2006) and an extracellular proteoglycan (Delgado *et al.*, 2006). Thus it is not surprising that for many years the field has been divided as to whether MeCP2 acts as a repressor or an activator (Ben-Shachar *et al.*, 2009; Chahrour *et al.*, 2008).

Recently the relationship between MeCP2 binding, DNA methylation and gene expression changes by RNA-sequencing analysis has been analysed in order to reveal trends in individual datasets rather than looking for groups of commonly misregulated genes between datasets. Such analyses in mouse hypothalamus (Chen *et al.*, 2015), cortex and cerebellum (Gabel *et al.*, 2015; Kinde *et al.*, 2016) demonstrates that genes that are up-regulated in MeCP2 KO brains tend to have high levels of gene-body mCpA (mCpH in Chen study). As discussed in Chapter 1.2.4, MeCP2 ChIP-sequencing in these brain regions shows enrichment for genes that have high levels of gene-body mCpA, and that EMSA analysis and ChIP-seq normalised reads determines mCpG to be the strongest predictor of MeCP2 binding, followed by mCpH and then unmethylated DNA (Chen *et al.*, 2015). In line with this, genes that are repressed in MeCP2 transgenic mice tend to have higher levels of gene-body mCpH compared to the rest of the genome, suggesting mCpH is directing MeCP2 binding (Chen *et al.*, 2015). The requirement for mCpA as being a predictor of genes that

will be up-regulated in MeCP2 KO samples was observed in 3 KO brain regions, in R306C mice and in RTT patient brain samples (Gabel *et al.*, 2015) and was narrowed to mCAC recently (Lagger *et al.*, 2017). Furthermore *Dnmt3A* conditional KO mice (nestin-cre) demonstrate a loss of mCpH levels but no change in mCpG levels and microarray analysis in the cerebellum reveals a mCpA-dependent increase in gene expression levels in these mice (Gabel *et al.*, 2015). But, MeCP2 binding is not abolished in these mice, demonstrating that mCpG dinucleotides are still important for MeCP2 occupancy (Kinde *et al.*, 2016). A direct comparison of gene expression changes in the MeCP2 null and *dnmt3a* null mice would be useful for distinguishing the contributions of mCpG and mCpA binding for MeCP2-mediated gene repression. While the Chen study found genes that were down-regulated in the absence of MeCP2 (MeCP2-activated genes) were also enriched for gene-body mCpA in the hypothalamus, this finding was not recapitulated in analysis of the same dataset as well as those from cortex and cerebellum (Kinde *et al.*, 2016; Lagger *et al.*, 2017). In fact, genes that are repressed in the absence of MeCP2 do not seem to share any distinguishable features with regards to mCpG, mCpA, gene length or histone acetylation status suggesting that these genes are not direct targets for MeCP2-dependent regulation, but are instead secondary transcriptional effects (Kinde *et al.*, 2016; Lagger *et al.*, 2017).

To conclude, the current evidence suggests that MeCP2 acts as a transcriptional repressor, via the action of binding to sites of mCpG and mCAC in order to temper gene expression. As opposed to functioning like a flicker switch and turning gene expression off by direct binding, MeCP2 seems to act like a dimmer switch to dampen gene expression and provide an extra level of transcriptional constraint. Gene expression changes are not large upon the loss of MeCP2, but they are numerous and they consistently correlate with genes that have high levels of gene-body methylation. This suggests that MeCP2 functions to restrain gene expression via binding within the body of genes. How MeCP2 achieves this repression mechanistically is still unknown.

1.2.6 Other suggested roles for MeCP2 function

While there is strong evidence for the role of MeCP2 in transcriptional repression, multiple other models have been proposed, reviewed in (Lyst & Bird, 2015) (Figure 1C). The relevance of some of these functions for RTT pathology are yet to be established, but it is possible that through interactions with a number of different proteins MeCP2 acts as multi-functional hub.

Transcriptional Activation

In the hypothalamus of MeCP2 null mice, more genes are downregulated then upregulated and the reverse is true for mice that overexpress MeCP2 (Chahrour *et al.*, 2008). A subset of genes showed this reciprocal effect suggesting that MeCP2 could be activating gene transcription in the hypothalamus. Similar results have also been found in the cerebellum (Ben-Shachar *et al.*, 2009) and a mechanism for how these transcriptional changes take place has been put forward. Specifically, an interaction between MeCP2 and CREB (cAMP-responsive element binding protein 1) which is a known transcriptional activator has been suggested (Chahrour *et al.*, 2008). In line with an activation hypothesis, MeCP2 has been observed to bind to unmethylated promoters in SH-SY5Y cells (Yasui *et al.*, 2007) and in olfactory neurons *in vivo* (Rube *et al.*, 2016). Furthermore, there is evidence in mouse and human ESC-derived neurons that total transcription is reduced in MeCP2 KO samples (Li *et al.*, 2013c; Yazdani *et al.*, 2012). MeCP2 has also been claimed to interact with the euchromatin-associated transcriptional activator Brahma (Harikrishnan *et al.*, 2005), however this interaction was later questioned and thus the only current hypothesis for how MeCP2 might activate transcription is through a potential association with CREB (Hu *et al.*, 2006).

miRNA processing

MeCP2 has also been implicated in the regulation of miRNA processing by a number of studies (Cheng *et al.*, 2014; Szulwach *et al.*, 2010; Tsujimura *et al.*, 2015; Urdinguio *et al.*, 2010; Wu *et al.*, 2010). The relevance of these studies is uncertain, however, as Cheng and colleagues claim MeCP2 inhibits miRNA processing by interfering with the Drosha complex whereas Tsujimura and colleagues claim MeCP2 acts as part of the Drosha complex to facilitate miRNA processing. Furthermore, the region of MeCP2 that has been mapped as the Drosha interacting domain (the C-terminus of MeCP2) is indispensable for MeCP2 function because mice that are truncated at residue 312 are phenotypically WT (Sabine Lagger *et al*, unpublished results).

Regulation of splicing

MeCP2 has been implicated in the regulation of splicing via an RNA-dependent interaction with YB-1 (Young *et al.*, 2005). A YB1-interaction domain has been mapped to residues 195-310 and alternative splicing was found misregulated in mice that express a truncated form of MeCP2 (Shahbazian *et al.*, 2002b). Furthermore, two papers have been published this year reigniting the link between MeCP2 and splicing. Cheng and colleagues found that

MeCP2 affects the use of alternative exons by binding to the 5hmC DNA modification (Cheng *et al.*, 2017) and while MeCP2 has been shown to be able to bind to this modification by other laboratories (Gabel *et al.*, 2015; Lagger *et al.*, 2017; Mellen *et al.*, 2012; Valinluck *et al.*, 2004), it exists at such a low level in the adult mouse and human brain that its biological relevance is questionable (Lister *et al.*, 2013). The second paper proposed a mechanism whereby reduced DNA methylation results in reduced MeCP2 binding throughout gene bodies leading to reduced splicing and increased intron retention in primary mouse granulocytes (Wong *et al.*, 2017). A similar observation has been seen in human lung and colon carcinoma cell lines (Maunakea *et al.*, 2013), however the relevance of these findings for the *in vivo* function of MeCP2 in the CNS and for RTT disease pathology is unknown.

Chromatin compaction/remodelling

Several studies have demonstrated a role for MeCP2 in chromatin remodelling and chromatin interactions. Early *in vitro* experiments demonstrated the ability of MeCP2 to bind to nucleosomal plasmid DNA and compete with H1 (Nan *et al.*, 1997). MeCP2 was also shown to bind to nucleosomal linker DNA both *in vitro* and *in vivo* (Ishibashi *et al.*, 2008; Nikitina *et al.*, 2007a) and is able to compete for binding with H1 in mouse fibroblasts, being more effective at displacing H1 than *vice versa* (Ghosh *et al.*, 2010). In support of a mechanism whereby MeCP2 preferentially binds to methylated cytosines in linker DNA, thus displacing H1, H1 levels are surprisingly low in neurons (Pearson *et al.*, 1984), and were found to be increased 2-fold in MeCP2 KO neuronal nuclei (Skene *et al.*, 2010). In addition, a number of studies have shown that MeCP2 can induce chromatin compaction. Electron microscopy using nucleosomal arrays demonstrated how MeCP2 is able to bind arrays and cause compaction (Georgel *et al.*, 2003; Nikitina *et al.*, 2007b). In addition, looping of DNA has been observed both *in vitro* (Ghosh *et al.*, 2010) and *in vivo* (Horike *et al.*, 2005). Specifically, atomic force microscopy observed a single MeCP2 molecule making numerous contacts with a single strand of DNA thus creating observable DNA loops (Ghosh *et al.*, 2010), while a specific loop at the *Dlx5-Dlx6* imprinted locus in mouse brain is lost upon MeCP2 KO (Horike *et al.*, 2005). The biological significance of this is questionable as a second group could not replicate the MeCP2-loop-dependent silencing of this locus (Schule *et al.*, 2007). More recently array tomography microscopy has revealed alterations in the chromatin landscape in MeCP2 KO cells in brain slices from a *Mecp2*$^{+/-}$ female mouse (Linhoff *et al.*, 2015). If MeCP2 was involved in chromatin compaction one would expect an increase in the volume of heterochromatin in the absence of MeCP2, whereas the opposite

was found. Heterochromatic foci volumes were increased in nuclei that lacked MeCP2 but a larger quantity of DAPI staining was present within these foci, indicating a redistribution of DNA into larger heterochromatic foci. Interestingly, Linhoff and colleagues observed increased chromatin compaction in hippocampal CA1 pyramidal neurons and dentate granule neurons, but not in cerebellar granule neurons, which corresponds with the lower expression level of MeCP2 in the cerebellum (Ross *et al.*, 2016).

Inhibitor of retrotransposon

Retrotransposons are interspersed, repetitive elements within the mammalian genome, which can be transcribed and often retain the ability to transpose into new locations within the genome. The highly methylated state of their integrated form suggests that MeCP2 could play a role in the regulation of transcription and transposition of these elements. In line with this hypothesis, exogenous expression of MeCP2 (but not MBD1 or MBD2) in HeLa cells reduces the frequency of L1 retrotransposition, thus suggesting that MeCP2 inhibits the transposition of repetitive elements (Yu *et al.*, 2001). Furthermore, increased transposition of the L1 retrotransposon has been observed in the brains of *Mecp2*$^{-/y}$ KO mice and neural precursors derived from RTT patient iPSCs (Muotri *et al.*, 2010) and transcripts derived from repetitive elements are somewhat increased in NeuN+ nuclei from *Mecp2*$^{-/y}$ brain (Skene *et al.*, 2010) and in the brains of *Mecp2*$^{-/y}$ KO mice (Muotri *et al.*, 2010). The relevance of this phenotype remains to be determined, but the fact that RTT-like phenotypes in mice can be reversed upon re-expression of MeCP2 suggests that the irreversible insertional mutagenesis imposed by retrotransposition is not a major contributor (Guy *et al.*, 2007). MeCP2-induced repression of retrotransposon expression, however, could still be relevant for RTT pathologies.

Non-methylated DNA binding

While MeCP2 can specifically recognise and bind to sites of methylation in the genome there is evidence that this methylation-specific binding is assisted by non-methylated DNA recognition. In particular, the observation that depletion of DNA methylation does not result in the complete dissociation of MeCP2 from DNA but rather its redistribution suggests that MeCP2 has the ability to bind to non-methylated sites *in vivo* (Baubec *et al.*, 2013). *In vitro* data has shown that MeCP2 can bind to nucleosomal arrays regardless of the methylation status of the underlying DNA, and that this non-specific binding may be modulated by the C-terminus of MeCP2 (Georgel *et al.*, 2003; Ghosh *et al.*, 2010; Nikitina *et al.*, 2007b). This *in vitro* data is supported by *in vivo* ChIP analysis in SH-SY5Y human neurons (Yasui *et al.*,

2007), mouse cortical neuronal cultures (Harikrishnan *et al.*, 2010) and mouse olfactory neurons (Rube *et al.*, 2016) showing binding of MeCP2 to non-methylated loci.

Three outstanding questions remain though; what domains within MeCP2 specifically mediate this non-methylated DNA binding, to what extent does this binding occur genome-wide, and how critical is this mode of DNA binding for Rett syndrome? To address the former question, there has been some focus on the three potential AT-hook motifs within MeCP2 that, if functional, would mediate AT-rich sequence binding of MeCP2 (Figure 1). The most C-terminal AT-hook (AT-hook 3) is not recognised as an AT-hook motif by *in silico* analysis and in fact has sequence disruptions in the core motif suggesting that it cannot function as an AT-hook. The other two AT-hooks have recently been assessed *in vitro* and *in vivo* for AT-hook character, and while AT-hook 1 can mediate AT-rich DNA binding in the absence of a functional MBD, AT-hook 2 is less robust (Lyst *et al.*, 2016). Truncation of MeCP2 at arginine 168 results in decreased binding of MeCP2 to non-methylated nucleosomal arrays *in vitro*, although does not abolish it (Georgel *et al.*, 2003). Alternative truncations to R274X and H370X also dramatically reduce DNA binding and abolish it when R274X is combined with the MBD-containing point mutation R106W, thus suggesting that the C-terminal domain is critical for non-methylated DNA binding (Nikitina *et al.*, 2007b). In addition, the MBD, the intervening domain and the TRD have all been implicated in non-methylated DNA binding as well [reviewed in (Hansen *et al.*, 2010)].

Perhaps all of these domains contribute to non-methylated DNA binding of MeCP2, implicating multi-modal DNA binding properties. MeCP2 may bind non-discriminately throughout the genome, but have higher affinities (and perhaps longer residence times) in the presence of methylated cytosines. *In vivo* ChIP analysis of the MeCP2 truncations and point mutations used in these *in vitro* experiments will help to determine the ability of these proteins to bind to non-methylated DNA in an endogenous context. In addition, downstream mouse phenotyping or neuronal transcriptome analysis will help to determine the significance of this mode of binding for proper MeCP2 function.

1.3 Rett syndrome and MeCP2-related disorders

In 1999 mutations in MeCP2 were discovered as the leading cause of Rett syndrome (RTT: Online Mendelian Inheritance in Man #312750), a severe neurological disorder and one of the leading causes of mental retardation in females (Amir *et al.*, 1999). RTT was first

described in 1966 by Andreas Rett, an Austrian clinician who reported his observations on 22 girls who had similar clinical features (Rett, 2016). It was not until 1983, however, when Bengt Hagberg published his own observations of this disease that the term Rett syndrome was coined (Hagberg *et al.*, 1983). RTT has a prevalence of 1 in every 10000 live female births and due to the X-linked nature of its genetics, primarily affects females. Despite early predictions that RTT was a disorder of the X chromosome, it took many years for the genetic cause to be identified due to the sporadic nature of the disorder (Hagberg *et al.*, 1983). With about 12 familial cases of RTT having been identified worldwide (i.e. multiple cases of RTT occurring within the same family tree), exclusion mapping using microsatellite markers on the X chromosome led to the localisation of the genetic cause to the Xq28 region on the X chromosome (Curtis *et al.*, 1993; Schanen *et al.*, 1997; Sirianni *et al.*, 1998). Genetic screening of genes located within this region eventually led to the identification of *MECP2* as the causative agent (Amir *et al.*, 1999). Women and girls with the disorder are heterozygous for the MeCP2 mutation and exhibit mosaicism whereby ~50% of their cells express a wild-type (WT) copy of *MECP2* and the other ~50% express a mutant version. Skewed X chromosome inactivation (XCI) can therefore result in patients with the same *MECP2* mutation displaying varied clinical severity. Hemizygous males, having only one X chromosome are more severely affected by *MECP2* mutations, often die prenatally and those that are born display severe encephalopathy (Katz *et al.*, 2016; Schanen *et al.*, 1998). Genetic screening is now a routine step as part of the diagnostic criteria for RTT, and 660 different *MECP2* mutations have been identified and are documented in RettBase (http://mecp2.chw.edu.au/) (Christodoulou & Ho, 2003). There are eight recurrent missense or nonsense mutations that account for approximately 2/3 of RTT mutations (Leonard *et al.*, 2017), all of which disrupt the MBD or NID of MeCP2, thus highlighting the importance of these two domains for MeCP2 function (Lyst & Bird, 2015) (Figure 1B).

In addition to Rett syndrome, another MeCP2 related disorder has recently been characterised called MeCP2 duplication syndrome (Van Esch *et al.*, 2005). It is caused by duplications of portions of the X chromosome specifically including *MECP2* and *IRAK1*, but duplication size and breakpoint locations vary in the population. Due to the duplication occurring most often *in cis* onto the same X chromosome, the condition affects boys while females are protected by random XCI. The frequency of MeCP2 duplication syndrome has not been extensively studied as it is a newly recognised disorder and very few cases have been reported, however it is predicted to account for 1% of X-linked mental retardation cases (Ramocki *et al.*, 2010). This X-linked neurodevelopmental syndrome is characterised by

developmental delay, severe to profound mental retardation, absent speech, progressive spasticity and epilepsy (Ramocki *et al.*, 2010). Two mouse models exist, both of which were generated prior to the condition being defined. The first model knocked a *Mecp2* cDNA construct into the *Tau* locus thus providing neuron-specific MeCP2 overexpression (Luikenhuis *et al.*, 2004), while the 2nd model consists of a 99 kb human PAC containing *MECP2* driven by its own promoter (Collins *et al.*, 2004). Both mouse models exhibit similar phenotypes suggesting neuronal MeCP2 overexpression is the major driver of disease progression. Phenotypes include growth retardation, progressive neurological symptoms and ataxia. The two domains that are implicated in Rett syndrome pathologies (the MBD and the NID) are also implicated in MeCP2 duplication syndrome pathologies as mice that overexpress MeCP2 containing MBD or NID mutations do not develop phenotypes (Heckman *et al.*, 2014; Lamonica *et al.*, 2017). Likewise, overexpression of WT MeCP2 in a RTT mutation background can prevent the onset of RTT-like phenotypes in mice (Luikenhuis *et al.*, 2004).

1.3.1 Clinical manifestations of RTT

RTT is characterised by a normal period of development for the first 6-18 months of life, followed by developmental stagnation, developmental delay and subsequent developmental regression, all within the first 3 years of life (Kaufmann *et al.*, 2016). Loss of motor skills, loss of verbal communication and the replacement of purposeful hand movements with hand stereotypies and hand mouthing are hallmarks of RTT. Approximately half the girls can walk, however often with an abnormal gait. Microcephaly, seizures, scoliosis, apneas, hyperventilation and cardiac problems are all other features of RTT. During the regression phase of the disease, girls often develop autism-like behaviours, and while this can continue for a minority of patients, RTT is not classified as an autism spectrum disorder (ASD) (Katz *et al.*, 2016; Neul *et al.*, 2010). Often a 2nd period of regression can occur at the late adolescent stage whereby patients develop Parkinsonian features (Humphreys & Barrowman, 2016). Despite these severe symptoms from an early age, women with RTT who have the necessary care can survive well in adulthood and the survival rate for the disorder at age 45 is over 70% (Tarquinio *et al.*, 2015).

1.3.2 Neurobiology of RTT

Post-mortem analyses of brains from RTT patients have revealed a number of morphological phenotypes. At a gross anatomical level, the brains of RTT patients are smaller, and contain

less melanin than age-matched control brains (Armstrong, 2002). When this is looked at in closer detail, individual neurons are smaller, have shorter neurites, increased cell packing, reduced dendritic complexity and fewer spines (Dawna Armstrong, 2005; Dawna Armstrong *et al.*, 1995; Dunn *et al.*, 2014; Kaufmann *et al.*, 2000; M.L. Bauman *et al.*, 1995). While these malformations are found throughout the entire brain, the extent varies from region to region. For example, the cerebral hemispheres have a more pronounced decrease in volume than the cerebellum, which could perhaps reflect the smaller level of expression of MeCP2 in the cerebellum compared to other brain regions (Dawna Armstrong, 2005; Ross *et al.*, 2016). Other malformations that have been reported include decreased cortical thickness, decreased MAP2 expression, decreased dopamine in the hippocampus, increased cell death in the subtania nigra pars compacta and decreased choline acetyltransferase in the forebrain (Colantuoni *et al.*, 2001; Kaufmann *et al.*, 2000; Kitt & Wilcox, 1995).

While MeCP2 is ubiquitously expressed, its high levels in the brain and the primarily neurological symptoms observed in patients with Rett syndrome suggested that the central nervous system plays a crucial role in the disease. The first mouse models of RTT immediately highlighted this by the use of Cre-loxP technologies to delete MeCP2 from the central nervous system using *nestin*-cre (deletes in neurons and glia from embryonic day 12) (Chen *et al.*, 2001; Guy *et al.*, 2001). These studies found a near identical phenotype in these mice compared to the ubiquitous-cre KO mice, highlighting the central nervous system as being the major contributor to the RTT-like phenotype in mice (Guy *et al.*, 2001). Post-natal KO of MeCP2 in post-mitotic neurons using *CamK*-cre that is expressed from P21 revealed that mice were healthy for the first 3 months of life, before developing RTT-like phenotypes (Chen *et al.*, 2001). *CamK*-cre expression deleted MeCP2 from the forebrain, hippocampus and brainstem but not from the cerebellum. Interestingly, post-mortem examination of cell soma sizes revealed decreases in the cortex and hippocampus but not in the cerebellum, highlighting a direct link between MeCP2 deficiency and neuronal soma size in mouse models, which recapitulates the phenotype seen in patients (Chen *et al.*, 2001). Of note, the *CamK*-cre mice did not delete MeCP2 from glia, and while the phenotype was similar to that of germline-null mice, it was less severe. Whether this absence of a severe phenotype is due to MeCP2 expression in glia, in the cerebellum or due to the delayed KO of MeCP2 is unknown. Of note, however, mice that have a glial-specific KO of MeCP2 using *hGFAP2*-cre display normal lifespan, locomoter functions, anxiety and dendritic branching but exhibit hindlimb clasping and irregular breathing, potentially suggesting that the breathing phenotype in mice and RTT patients is due to a loss of MeCP2 in both neurons and glial

populations. Performing the reverse experiment, however, of rescuing null mice with glial-specific expression does rescue a number of overt RTT-like phenotypes, including breathing irregularities and dendritic branching (Lioy *et al.*, 2011).

Through conditional KO studies (cKO) MeCP2 has also been deleted from a range of different neuronal sub-types and different brain regions in order to understand the contributions of different neurons to disease pathology, reviewed in (Guy *et al.*, 2011). Perhaps unsurprisingly, the phenotypes observed in the mice reflect the function of the brain region from which MeCP2 was deleted. For example, deletion of MeCP2 specifically in areas of the hypothalamus using *Sim1*-cre results in overweight mice that suffer from hyperphagia, but do not have any motor, learning or memory deficits (Fyffe *et al.*, 2008). Alternatively, deletion in dopaminergic and noradrenergic neurons using *Th*-cre results in hypoactivity and abnormal motor coordination but normal anxiety levels (Samaco *et al.*, 2009). Similar results have been found in mouse models using *Pet1*-cre (serotonergic neurons) (Samaco *et al.*, 2009), *Viaat*-cre (GABAergic neurons) (Chao *et al.*, 2010), *Dlx5/6*-cre (forebrain GABAergic neurons) (Chao *et al.*, 2010; Zhang *et al.*, 2014), *Emx1*-cre (forebrain excitatory neurons) (Zhang *et al.*, 2014), *PV*-cre (parvalbumin neurons) (Ito-Ishida *et al.*, 2015), *SOM*-cre (somatostatin neurons) (Ito-Ishida *et al.*, 2015) and *Vglut2*-cre (glutamatergic neurons) (Meng *et al.*, 2016). These studies have demonstrated the importance of MeCP2 in all neuronal sub-types and brain regions. The only cKO mouse that has a decreased lifespan is the *Viaat*-cre line (50% survival at 26 weeks compared to ~8 weeks for MeCP2 null mice) suggesting that perhaps the loss of MeCP2 in GABAergic neurons is more critical for RTT pathologies. In line with this, rescue of MeCP2 expression in null mice specifically in GABAergic neurons is able to rescue lifespan, body weight, ataxia and apraxia (Ure *et al.*, 2016). Interestingly, while acoustic startle responses and tremors are not rescued in this mouse line, these phenotypes are rescued in a glutamatergic neuron-specific rescue, which itself also has an improved lifespan but does not rescue motor deficits (Meng *et al.*, 2016).

RTT is not considered to be a neurodevelopmental disorder due to the observation that reactivation of MeCP2 expression in null mice after symptoms have appeared can reverse the neurological phenotype in the mice (Guy *et al.*, 2007). This suggests that the neurological defects caused by loss of MeCP2 function are reversible and therefore MeCP2 is required for neuronal maintenance but not neuronal development. At the same time, MeCP2 is required for proper neuronal function at all stages of life, as deletion of MeCP2 at 3 weeks, 8 weeks,

11 weeks or 20 weeks leads to the onset of RTT-like phenotypes and premature death in male mice (Cheval *et al.*, 2012; McGraw *et al.*, 2011).

1.3.3 Therapeutic strategies for RTT

The discovery in 1999 that the majority of Rett syndrome cases are caused by mutations in the X-linked gene *MECP2* created a new era in the RTT field, often referred to as the post-MeCP2 era (Amir *et al.*, 1999). In the pre-MeCP2 era, most clinical trials were based on clinical observations, for example the observation that RTT patients had low L-carnitine plasma levels led to the clinical trial of L-carnitine which demonstrated some improvements in some patients (Ellaway *et al.*, 1999).

Upon the discovery of MeCP2 being the causative agent in Rett syndrome, and knowledge of its links to DNA methylation, trials were begun using methyl donors such as folane/betaine and creatine. Furthermore, with molecular details emerging that RTT is a disorder of the whole brain as opposed to a particular region, clinical trials using pan-specific growth factors such as BDNF and IGF-1 in order to restore the functionality of multiple neuronal circuits is being assessed as a therapeutic option. At least four clinical trials are ongoing or have been completed using this approach. Similarly, restoration of neurotransmitter balance is being assessed, however knowledge of which neurotransmitters to use, when in the disease progression to begin treatment and which patients will benefit from that particular treatment is currently lacking [reviewed in: (Kaufmann *et al.*, 2016)].

In the post-MeCP2 era, more targeted therapeutic strategies have been explored including direct manipulation of *MECP2* itself. The key study came in 2007 when restoration of MeCP2 in *Mecp2*-deficient mice led to dramatic restoration of breathing, motor skills and other RTT-like phenotypes in the mice (Guy *et al.*, 2007). This strategy provided the conclusive evidence that RTT is a reversible disorder in mice, and is therefore potentially curable in patients, even after the onset of symptoms. Since then, MeCP2-targeted therapeutic strategies have included replacement of MeCP2 via gene or protein therapy, reactivation of the inactive *MECP2* allele and the use of read-through compounds for those patients who have a mutation that produces an early stop codon (Katz *et al.*, 2016; Shah & Bird, 2017).

Gene therapy, i.e. providing a WT copy of the *MECP2* gene, has so far been the most explored therapeutic avenue. Two labs have shown in mouse models that injections of

adeno-associated viruses (AAV) containing *Mecp2* reverse some of the RTT-like features in mice (Gadalla *et al.*, 2013; Garg *et al.*, 2013). AAV vectors for gene therapy are widely used for a number of disorders and in some cases the results are very promising, for example delivery of *rpe65* to the sub-retinal layers in the eye to reverse the blindness observed in *rpe65* deficiency [reviewed in: (Samulski & Muzyczka, 2014)]. Indeed, AAV vectors are the first example of viruses being approved for use in the clinic in the form of Glybera, an AAV virus to treat lipoprotein lipase deficiency (Bryant *et al.*, 2013). While AAV vectors seem extremely promising, there are some worries over the use of this treatment for Rett syndrome. In particular is the risk of curing Rett syndrome, but simultaneously causing MeCP2 duplication syndrome, while another issue is inducing liver toxicity. To circumvent these potential problems, alternative delivery routes and a range of doses are being tested for safety and efficacy in mouse models. Most recently, cerebroventricular injections of 1×10^{11} vector genomes/mouse were shown to be most effective at achieving phenotypic rescue and avoiding liver toxicity (Gadalla *et al.*, 2017). Tight control over the expression of MeCP2 would also be desirable, but the small packaging size of the AAV genome (~4.7kb) limits the possibilities here. Small versions of the endogenous *Mecp2* promoter have proven to be successful at controlling endogenous expression levels of MeCP2 (Gadalla *et al.*, 2017; Rastegar *et al.*, 2009), and the use of smaller, unstable, partially impaired versions of MeCP2 looks promising too (Lamonica *et al.*, 2017)(Tillotson *et al.*, under review). If reproducible dosage regimes can be established through the use of inferior MeCP2 molecules (i.e. unstable or partially functional), endogenous promoters and other regulatory elements, there is still the concern that the endogenous mutant protein will interfere with the function of the exogenously applied WT protein. Recent experiments in mice modelling MeCP2 overexpression syndrome have suggested that overexpression of T158M, R111G and R306C MeCP2 protein in WT male and female mice do not produce any dominant negative effects, suggesting that at least for these mutations, this may not be a problem (Heckman *et al.*, 2014; Lamonica *et al.*, 2017).

An alternative to gene therapy is protein therapy, i.e. restoring the levels of WT MeCP2 protein. This could involve the use of read-through compounds to encourage RNA polymerase to read-through early stop codons in *MECP2*, as in the case of R168X (Brendel *et al.*, 2011). While this has been successful in cultured mouse cells, this has not been translated to *in vivo* systems and, like the case for reactivation of the inactive X chromosome allele, specificity of targeting *MECP2* alone is currently limited [reviewed in (Katz *et al.*, 2016)]. Delivery of MeCP2 protein itself is an alternative and more novel strategy being

explored because the chemically basic nature of MeCP2 itself makes it an excellent protein candidate for traversing the cellular membrane (discussed in chapter 7) (Kabouridis, 2003; Nan *et al.*, 1996). While a patent has been taken out in the United States for the delivery of MeCP2 protein as a therapeutic strategy for Rett syndrome (Laccone, 2012), evidence for its effectiveness at transducing cells, and in particular neurons, is limiting. Experiments to address this will be described at the end of this study in chapter 7.

1.3.4 Current systems for studying MeCP2 and RTT

Mouse models have been an invaluable tool for advancing our knowledge of the function of MeCP2 and its role in Rett syndrome (for a full review see (Guy *et al.*, 2011)). The first mouse models were reported in 2001 when two separate laboratories created MeCP2-null mouse lines via deletion of exon 3 (Chen *et al.*, 2001) and deletion of exons 3 and 4 (Guy *et al.*, 2001). Both lines exhibited multiple neurological and motor phenotypes reflective of the disease symptoms, including smaller brains, abnormal gait, and breathing abnormalities. While the heterozygous female mice are the true RTT model, hemizygous male mice are more commonly studied due to the strong phenotype, early death and faster onset of symptoms. Furthermore, direct genotype-phenotype correlations can be drawn more easily in hemizygous males than they can in mosaic heterozygous females. The observation that a central nervous system (CNS)-specific KO of MeCP2 in mice produced extremely similar phenotypes to the complete null, led to the generation of multiple brain region-specific KOs which were described in Chapter 1.3.2. More recently, KO in peripheral tissues (PKO) further confirmed the loss of MeCP2 function in the CNS as being the major contributor to RTT-like phenotypes (Ross *et al.*, 2016). Using *nestin*-cre to remove a floxed STOP cassette in the *Mecp2* reading frame in mice, expression of MeCP2 could be restored to central nervous system tissues while leaving peripheral tissues devoid of MeCP2 expression. The only phenotypes in the mice were hypoactivity and decreased bone stiffness, demonstrating the importance of MeCP2 expression in the CNS as the primary origin of RTT-like phenotypes (Ross *et al.*, 2016).

In addition to *Mecp2*-null mouse models, there have been a number of lines containing RTT patient mutations, the more accurate model system. The first of these was the knock-in (KI) of a stop codon at position arginine 168 (R168X) which exhibit the same phenotype as null mice lines (Lawson-Yuen *et al.*, 2007). Other RTT nonsense mutation models include R255X (Pitcher *et al.*, 2015), R270X-EGFP and R273X-EGFP (Baker *et al.*, 2013). Like R168X the R255X mutation was knocked into the endogenous *Mecp2* locus. Zero protein

was observed resulting in a phenotype extremely similar to the null mice (Pitcher *et al.*, 2015). On the other hand, the R270X-EGFP and R273X-EGFP alleles were introduced as transgenes in a null background and mouse lines were selected for based on the expression of the truncated protein at similar levels to WT protein in WT mice. Surprisingly the addition of 3 extra amino acids improves the survival of the mice from 12 weeks to 29 weeks and it was believed that the difference in severity was due to the retention of a full AT-hook motif in the G273X allele; referred to as AT-hook 2 in Figure 1 (Baker *et al.*, 2013). Recent *in vitro* and *in vivo* data directly assessing the AT-hook character of this sequence has cast doubt over this, finding that it has very weak AT-hook character (Lyst *et al.*, 2016). While the mice exhibited differences in lifespan with the two mutations, there is only 1 boy who has been reported to have a G273fs mutation and although he lived longer than patients with R270fs mutations, it is difficult to draw conclusions on the functional importance of these 3 amino acids based on one case (Ravn *et al.*, 2003). There is also one reported case of a female with a G273fs mutation but no clinical assessment has been published (Zahorakova *et al.*, 2007). Perhaps KI of the two mutations into the endogenous locus would be more informative, particularly if the precise clinically observed frameshift and nonsense mutations were modelled. There have also been KI or transgenic mice expressing R111G-EGFP, R133C-EGFP, T158M-EGFP, T158A, R306C and R306C-EGFP mutations (Brown *et al.*, 2016; Goffin *et al.*, 2012; Heckman *et al.*, 2014; Lamonica *et al.*, 2017; Lyst *et al.*, 2013). Interestingly the clinical severity of point mutations in patients is mirrored by the severity of these mutations in mice (Brown *et al.*, 2016). Two other mouse models have been produced that do not mimic known RTT mutations; a truncation of MeCP2 in which T308 is converted to a stop codon results in a delayed but overt phenotype in the mice (Shahbazian *et al.*, 2002b). The A140V mutation is not associated with RTT but is observed in several cases of X-linked mental retardation. Despite observations of increased cell packing and reduced dendritic arborisation in the brain, A140V mice showed no neurological symptoms and had a normal lifespan (Jentarra *et al.*, 2010).

Other animals models include *Xenopus laevis* (Stancheva *et al.*, 2003), a rat KO model (SAGE laboratories) and two monkey models (Chen *et al.*, 2017; Liu *et al.*, 2014a; Liu *et al.*, 2014b). Recent analysis of one of the monkey models revealed a number of phenotypes, including those not detected in rodent models, for example male embryonic lethality and social withdrawal, although at 20 months of age no developmental delays or growth delays were observable (Chen *et al.*, 2017). MeCP2 deficiency in rats results in numerous neurological phenotypes including decreased brain weight, reduced survival, hindlimb

clasping, general lethargy, abnormal gait, reduced exploratory behaviour, decreased locomotion and abnormal respiratory patterns including apneas, all of which are observed in the mouse models (Patterson *et al.*, 2016). In addition, female MeCP2 KO rats display acquirement of fine motor skills followed by regression of development thus mirroring the symptom development in RTT patients, a finding that has not yet been observed in mice (Veeraragavan *et al.*, 2016).

Finally, a number of neuronal tissue culture systems have been utilised in order to assess the molecular function of MeCP2 such as embryonic stem cells (ESCs) and induced pluripotent stem cells (iPSCs) from RTT patient fibroblasts. Patient iPSC-derived neuronal lines include a wide variety of missense, nonsense and framseshift mutations (Ananiev *et al.*, 2011; Cheung *et al.*, 2011; Kim *et al.*, 2011; Marchetto *et al.*, 2010; Tang *et al.*, 2016; Zhang *et al.*, 2016). In addition, MeCP2 has been knocked out in both mouse and human ESCs and the resulting neurons studied (Li *et al.*, 2013c; Yazdani *et al.*, 2012). Many novel phenotypic observations have been uncovered using such systems, such as reduced total RNA amount in null neurons, strong transcriptional phenotypes and electrophysiological defects which has broadened our understanding of the function of MeCP2. Details of experiments using these systems will be discussed in Chapter 5 and are detailed in Supplementary Tables 1 - 4.

Mouse studies have been invaluable for determining the physiological relevance of MeCP2 function in the CNS. Further studies in rats to study the social and fine cognitive defects in RTT models may be useful, particularly with relation to testing therapeutic strategies in rodent models (Shah & Bird, 2017). Combining model systems will enable a deeper understanding of the function of MeCP2 and will help to uncover the molecular details that underlie Rett syndrome pathologies. In particular, the use of human neurons in tissue culture combined with rodent models may prove to be a powerful pipeline for screening and discovery of therapeutics for RTT. Rescue of a particular molecular phenotype in tissue culture, combined with observation of improved symptomatic scores in rodents may prove to be indicative of a successful drug in clinical trials.

1.3.5 LUHMES cells as new model system

In the gene editing era, the use of human iPSCs as tissue culture models for probing the molecular and neuropathological details that underlie complex diseases is a common and useful approach to understanding neurological disorders. There are, however, some drawbacks to the use of these cells. Primarily, there can be a large variability in the quality of

iPS cell lines when they are derived which can impact the subsequent differentiation and phenotypic outcome of the experiment (Ardhanareeswaran *et al.*, 2017). Specifically, the variety of reprogramming methods that are used (Sendai virus, lentivirus, plasmid transfection, protein delivery) may impose some variability between experiments and thus make it difficult to compare between individual studies. The same can be said for the variety of neuronal differentiation protocols that are used in different laboratories (Broccoli *et al.*, 2015; Zhang *et al.*, 2013). In addition, a full reset of the epigenetic and transcriptional network in cells can sometimes be incomplete, leading to subtle differences between individual clonal cell lines (Ardhanareeswaran *et al.*, 2017; Choi *et al.*, 2015; Ohi *et al.*, 2011). This is particularly critical when studying Rett syndrome, where the disorder appears to be defined by small and subtle transcriptional changes. Alternative human neuronal cell lines include the SH-SY5Y cell line, CB660 and neural stem cells derived from foetal human brain (NSC StemPro). SH-SY5Y cells, however, are a neuroblastoma cell line and contain an abnormal karyosome (Krishna *et al.*, 2014), and neural stem cells take a long time to mature, expressing markers for neural precursors for 4 weeks (Shin & Vemuri, 2010; Sun *et al.*, 2008; Tong *et al.*, 2016).

The LUHMES cell line is an alternative and relatively new human neuronal cell line that will be used in this study. Mesencephalic cells from an 8-week old female foetus were immortalised using a retroviral transgene that constitutively expresses *v-myc* under the control of a tetracyline-controlled promoter (Hoshimaru *et al.*, 1996; Lotharius *et al.*, 2002). This keeps LUHMES cells in a proliferative, stem-like, but neuronal-committed state. Addition of tetracycline to the cell culture media switches off *v-myc* expression and LUHMES cells are therefore allowed to differentiate into mature, post-mitotic, dopaminergic neurons (Scholz *et al.*, 2011). This differentiation is extremely rapid and very robust, resulting in the generation of a population of neurons of a single sub-type within one week. So far, LUHMES-derived neurons have been used to model Parkinson's disease (Lotharius *et al.*, 2005; Xiang *et al.*, 2013), for cytotoxicity assays (Tong *et al.*, 2016) and for technology development (Dinh *et al.*, 2013; Hughes *et al.*, 2014; Ilieva *et al.*, 2013).

The benefits of LUHMES cells over other neuronal tissue culture systems are 3-fold. Firstly, the speed of differentiation is extremely advantageous. As differentiation begins from a neuronal precursor stage, mature, electrically active neurons expressing a wide complement of neuronal markers can be achieved in just 1 week (Figure 2) (Scholz *et al.*, 2011).

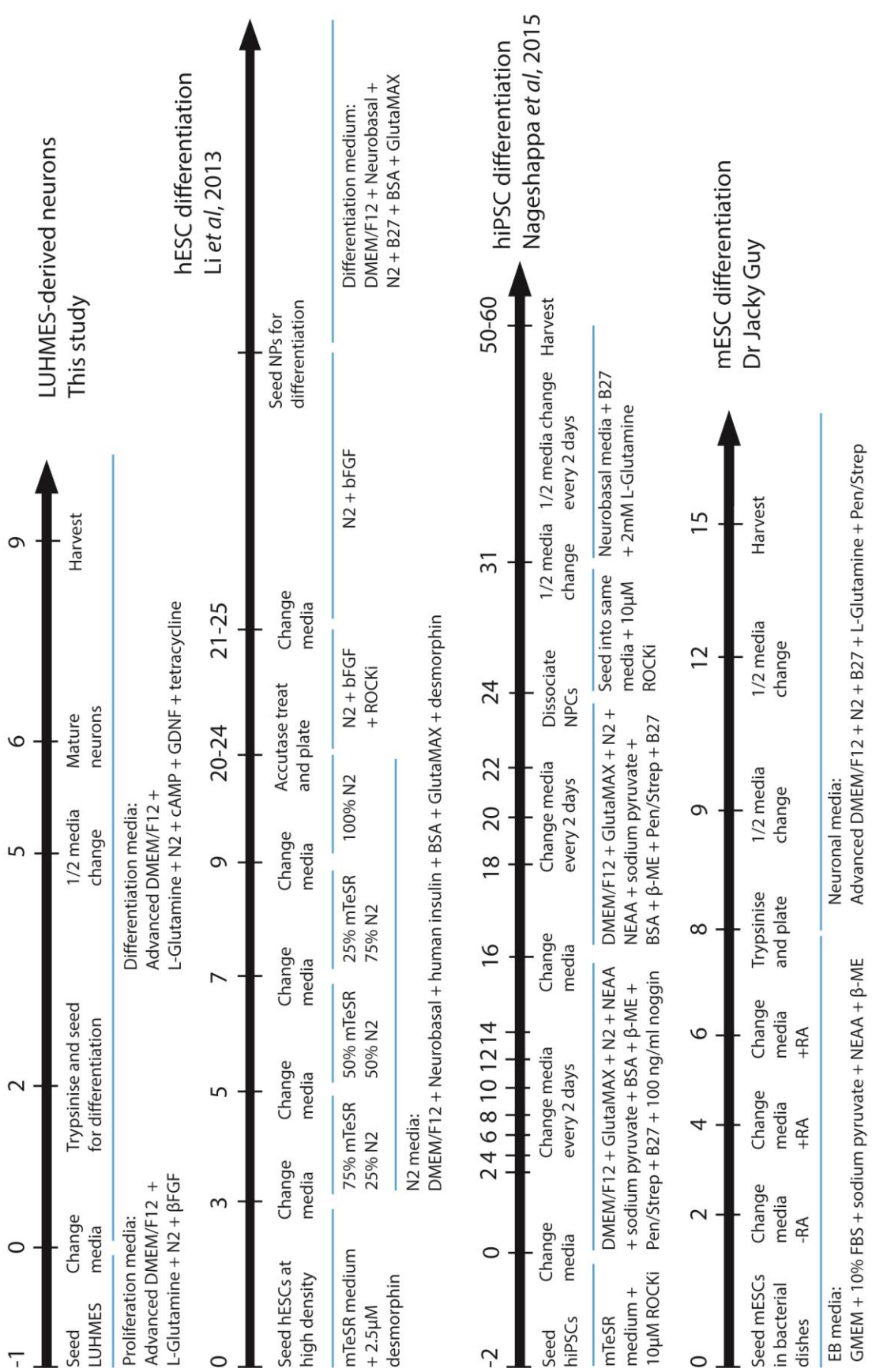

LUHMES-derived neurons
This study
-1
0
2
5
6
9
Seed LUHMES
Change media
Trypsinise and seed for differentiation
1/2 media change
Mature neurons
Harvest
Proliferation media:
Advanced DMEM/F12 +
L-Glutamine + N2 + βFGF
Differentiation media:
Advanced DMEM/F12 +
L-Glutamine + N2 + cAMP + GDNF + tetracycline
hESC differentiation
Li et al, 2013
0
3
5
7
9
20-24
21-25
Seed hESCs at high density
Change media
Change media
Change media
Change media
Accutase treat and plate
Change media
Seed NPs for differentiation
mTeSR medium + 2.5μM desmorphin
75% mTeSR
25% N2
50% mTeSR
50% N2
25% mTeSR
75% N2
100% N2
N2 + bFGF + ROCKi
N2 + bFGF
Differentiation medium:
DMEM/F12 + Neurobasal +
N2 + B27 + BSA + GlutaMAX
N2 media:
DMEM/F12 + Neurobasal + human insulin + BSA + GlutaMAX + desmorphin
hiPSC differentiation
Nageshappa et al, 2015
-2
0
2 4 6 8 10 12 14
16
18
20
22
24
31
50-60
Seed hiPSCs
Change media
Change media every 2 days
Change media
Change media every 2 days
Dissociate NPCs
1/2 media change
1/2 media change every 2 days
Harvest
mTeSR medium + 10μM ROCKi
DMEM/F12 + GlutaMAX + N2 + sodium pyruvate + BSA + β-ME + Pen/Strep + B27 + 100 ng/ml noggin
DMEM/F12 + GlutaMAX + N2 + NEAA + sodium pyruvate + BSA + β-ME + Pen/Strep + B27
Seed into same media + 10μM ROCKi
Neurobasal media + B27 + 2mM L-Glutamine
mESC differentiation
Dr Jacky Guy
0
2
4
6
8
9
12
15
Seed mESCs in bacterial dishes
Change media -RA
Change media +RA
Change media +RA
Trypsinise and plate
1/2 media change
1/2 media change
Harvest
EB media:
GMEM + 10% FBS + sodium pyruvate + NEAA + β-ME
Neuronal media:
Advanced DMEM/F12 + N2 + B27 + L-Glutamine + Pen/Strep

Figure 2. Tissue-culture neuronal differentiation timelines.

Comparison of differentiation time scales and media requirements for LUHMES cells, human embryonic stem cells, human induced pluripotent stem cells and mouse embryonic stem cells. Human ESC culture conditions taken from (Li *et al.,* 2013c), human iPSC conditions taken from (Nageshappa *et al.,* 2015), (Espuny-Camacho *et al.,* 2013) and (Gaspard *et al.,* 2009) and mouse ESC culture conditions were provided by Dr Jacky Guy.

Differentiation is not only quick, but is also rather simple. A single media change from β-FGF+ proliferative media to GDNF+, tetracycline+ differentiation media is all that is required to initiate differentiation. If required, cells can be passaged and counted two days after the media change providing a step to control for the number of neurons in the experiment; however, differentiation will proceed regardless of this step. Secondly, the homogeneous nature of the resulting population of neurons is extremely attractive. It has been proposed that gene expression studies using *Mecp2*$^{-/y}$ mouse brain samples are hampered by heterogeneous populations of different cell types which may mask significant gene expression changes of a single cell type. Studying a population of neurons, and indeed a population of a single subtype of neurons, may reveal robust and consistent effects that can be observed in population analysis experiments. Of note, current iPSC differentiation protocols are capable of producing populations of neurons with around 90% homogeneity when FACS-based enrichment approaches are used (Israel *et al.*, 2012) or by expression of neurogenin-2 in combination with astrocyte co-culture to produce excitatory neurons (Zhang *et al.*, 2013). Occasionally, however, these neurons can be a mix of neuronal subtypes, for example 15% of neurons being VGluT1-positive and 8% GABA-positive (Israel *et al.*, 2012). Thirdly, the fact that they are human cells may prove to be beneficial. LUHMES cells provide an easy to handle system within which to study the biology of endogenous MeCP2 in human neurons. There are numerous similarities between mouse and human samples, which is reflected in the 95% identical MeCP2 amino acid sequence between the two species and the strong phenotypic correlations when MeCP2 mutations are introduced (Brown *et al.*, 2016). Some of the more fundamental biology may, however, differ between the two species, for example different methods of MeCP2 protein regulation, different gene expression profiles, and different expression levels of MeCP2 partner proteins may exist.

On the other hand, iPSCs and NPCs offer the flexibility to differentiate into a variety of different neuronal subtypes as well as astrocytes (Chandrasekaran *et al.*, 2016; Dolmetsch & Geschwind, 2011; Juopperi *et al.*, 2012), whereas LUHMES cells are limited to a dopaminergic lineage (although the removal of cAMP and GDNF from the culture media can produce tyrosine hydroxylase negative neurons (Scholz *et al.*, 2011)). Due to this

heterogeneity in differentiation potential, hiPSCs can be used to develop 3D organoids for more complex analysis (Ardhanareeswaran *et al.*, 2017) and it remains to be seen if co-culture with astrocytes or 3-dimensional culture conditions will be suitable for LUHMES-derived neurons. Nevertheless, the other benefits of hiPSCs also apply to LUHMES cells; human genetics and proteomics, use of isogenic controls, high proportion of neurons in the sample and large sample sizes.

It is hoped that LUHMES cells will provide a simple, efficient and novel system within which to study MeCP2. In this study I will describe the use of gene editing techniques and a variety of experimental protocols in order to determine the utility of LUHMES-derived neurons for modelling and studying the molecular phenotypes involved in Rett syndrome. It is hoped that experiments using human neurons and a simplified, homogeneous model system will aid efforts towards deciphering the exact molecular function of MeCP2.

1.4 The CRISPR technology

The CRISPR locus was first observed in 1987 when Ishino and colleagues were sequencing the *iap* locus in *E. coli* and came across five highly homologous sequences of 29 nucleotides long that were separated by 32 nucleotide spacers (Ishino *et al.*, 1987). In 2000, another group studied the locus in a multitude of different bacteria and archaea and coined the term SRSRs (Short Regularly Spaced Repeats), although the function of these repeats was still unknown (Mojica *et al.*, 2000). In 2002 the name was changed to CRISPR which stands for Clustered Regularly Interspaced Short Palindromic Repeats and CRISPR-associated (Cas) proteins (Cas) were identified, two of which were predicted to function as a DNA helicase and as an exonuclease based on motif similarities to known proteins (Jansen *et al.*, 2002). *In silico* analyses found that the spacers in the CRISPR locus were homologous to bacteriophage DNA sequences and thus the hypothesis was put forward that the CRISPR locus functions as an RNA interference based immune response against invading bacteriophage (Bolotin *et al.*, 2005; Makarova *et al.*, 2006; Mojica *et al.*, 2005; Pourcel *et al.*, 2005). This hypothesis was confirmed in 2007, when Barrangou and colleagues noticed that upon bacteriophage challenge in *Streptococcus thermophilus* novel spacers that were homologous to portions of the bacteriophage genome were incorporated into the *S. thermophilus* genome at the CRISPR locus (Barrangou *et al.*, 2007). Furthermore deletion and addition of spacers into the CRISPR locus resulted in loss and gain of bacteriophage resistance respectively; and inactivation of two Cas genes in the CRISPR locus revealed the

importance of these proteins for bacteriophage resistance and for acquisition of novel spacers into the locus (Barrangou *et al.*, 2007).

The CRISPR system is composed of two main elements, all of which are transcribed from a single CRISPR locus. The first element is the long stretch of repeats and spacers that are transcribed as one long pre-mRNA transcript in the cell. This transcript is eventually processed into individual crRNAs (CRISPR RNAs), each one containing a single spacer element and therefore being able to target a single invading DNA sequence. The second element is the array of Cas proteins that recognise, process and bind to the crRNAs. There are multiple types of CRISPR systems across the 40% of bacteria and 90% of archaea that express this small-RNA based, adaptive immune system (Sorek *et al.*, 2008; Terns & Terns, 2011; Wiedenheft *et al.*, 2012).

1.4.1 CRISPR is a bacterial "immune system"

The ability of an organism to remember previous infections and viral challenges and to initiate an immune response based around this memory can be found in the cellular processes of RNA interference, PIWI RNAs and in the form of antibodies and lymphocytes. The discovery of an apparent immune memory that exists solely in bacteria and archaea therefore sparked much interest (Sorek *et al.*, 2008; Wiedenheft *et al.*, 2012).

While the components between different systems vary greatly, the overall mechanism of the "immune response" is conserved and consists of three stages: adaptation, crRNA biogenesis, and targeting of invader elements. In the adaptation stage, the acquisition of novel spacers into the CRISPR locus is achieved, and the more spacers derived from a single organism, the greater the "immune" response for the host cell. In the second stage the CRISPR locus is transcribed as one long pre-mRNA molecule that is processed to create individual crRNAs, which could be viewed as the equivalent of antibodies. For targeting, each crRNA forms a complex with Cas proteins and directs the Cas protein(s) (or CASCADE complex in some CRISPR systems) to the foreign genetic element via complementary base pairing. The Cas proteins then degrade the invading genetic material by introducing a double-strand break (DSB), and a successful immune response is achieved. The mechanistic details of the entire process in the different CRISPR systems is reviewed in Wright *et al.* (Wright *et al.*, 2016).

1.4.2 Adaptation of CRISPR for gene editing purposes

The discovery of an adaptive, directed, nuclease-based immune response was immediately recognised for its potential benefits in molecular biology and beyond. Providing phage resistance for economically important bacteria (for example in the food industry), being a tool for manipulation of nucleic acids, and limiting the spread of antibiotic resistance were among the first suggestions (Wiedenheft *et al.*, 2012). The adaptation of CRISPR systems for site-directed double-strand break (DSB) activity in heterologous cells was quickly established. Historically, targeted DSBs was achieved by meganucleases, zinc finger nucleases (ZFNs) and TAL effector nucleases (TALENs), all protein-based complexes that required complex cloning of a new protein every time a new locus wanted to be targeted for DSBs (Damian & Porteus, 2013). CRISPR, on-the-other-hand, had the potential to be a small RNA-based targeting technology, thus dramatically simplifying the cloning steps that would be needed for each new targeting experiment.

Two steps were required to repurpose CRISPR as a targeted DSB technology in heterologous cells. Identification of the minimal Cas proteins that are required to induce the DSB and identification of the minimal crRNA species necessary for direction of the Cas proteins to the targeted DNA locus. The Type II CRISPR system contains only 1 protein, Cas9, which is an endonuclease that binds to two RNA species; a crRNA and a tracrRNA (Charpentier & Doudna, 2013; Jinek *et al.*, 2012). The discovery that the two RNA molecules could be combined into a single guide RNA (sgRNA) was critical for repurposing the CRISPR system into a targeted DSB technology for use in mammalian cells (Jinek *et al.*, 2012). The first studies to take the CRISPR technology and utilise it in mammalian cells for targeted DSB purposes utilised the easy-to-transfect human embryonic kidney cell line, Hek293 cells (Cong *et al.*, 2013; Jinek *et al.*, 2013). Since then, CRISPR has been used to induce targeted DSBs in loci of interest in a plethora of cell lines and organisms including tobacco and rice (Jiang *et al.*, 2013b), miPSCs (Li *et al.*, 2016), hiPSCs (Mali *et al.*, 2013), *C. elegans* (Friedland *et al.*, 2013), *D. melanogaster* (Gratz *et al.*, 2013), Zebrafish (Hwang *et al.*, 2013), rats (Li *et al.*, 2013a; Li *et al.*, 2013b), pigs (Hai *et al.*, 2014) and monkeys (Niu *et al.*, 2014a). The generation of genetically modified mice has been performed by modifying the ESCs that are used to produce the mice or by direct injection of CRISPR molecules into the zygote (Wang *et al.*, 2013a). This technology has also been used to correct disease-causing mutations in mice as proof-of-principle studies for the use of CRISPR in the clinic, for example in the correction of cataracts (Wu *et al.*, 2013).

The most common mechanism for cellular repair of a double-strand break is to engage the non-homologous end joining (NHEJ) pathway which is inherently inaccurate and will result in the insertion and deletion of nucleic acids at the break point (Lieber, 2010). This non-precise process, however, has the advantage of inactivating the gene of interest and thus resulting in a functional gene knock-out (KO). In particular a gene KO will occur if an out-of-frame repair product is produced which results in an early stop codon in the reading frame and thus a truncated RNA product. If truncated early enough, this RNA will likely be degraded by nonsense mediated decay and result in a protein KO (Houseley & Tollervey, 2009). Alternatively, if the stop codon occurs later in the RNA molecule, a truncated (and potentially non-functional) version of the protein of interest may be produced. Another possibility is that the repair process will produce an in-frame repair product that maintains the reading frame of the RNA species and results in a mutated protein. If this mutation occurs in a critical portion of the protein of interest, this can be used to study the presence of a non-functional protein in the cell type of interest.

A less efficient, but often more desirable, repair outcome utilises homology-directed repair (HDR). In this instance a template is used as a donor molecule to encourage sequence-specific repair of the DSB. This type of repair is most frequently used during late S and G2 stages of the cell cycle when sister chromatids and homologous chromosomes are available for accurate HDR (Heyer *et al.*, 2010). Exogenous molecules, however, that have arms of homology surrounding the DSB can be used as an alternative donor molecule for HDR. Conventional targeting techniques involve using a large plasmid as a donor molecule, which contains many kilobases of DNA as the homology arms (Smithies *et al.*, 1985; Thomas & Capecchi, 1987). With the advent of CRISPR technologies, these arms can be significantly reduced in size without a loss of efficiency, and small single stranded oligodeoxynucleotides (ssODNs) have also been shown to be effective donor molecules for knocking in (KI) point mutations and small tags into a gene of interest (Ding *et al.*, 2013).

The level of HDR differs dramatically between different tissues and cell types (Miyaoka *et al.*, 2016), with mESCs exhibiting the highest levels and levels as low as 0.05-1.5% being reported in hiPSCs (Kime *et al.*, 2016). Efforts to improve efficiencies are being assessed from multiple angles, including the use of small molecules to increase HDR/NHEJ ratios and intelligent design of cut sites and donor molecules to encourage the use of specific repair pathways. Recent improvements will be discussed in Chapter 4.6 and are reviewed in (Komor *et al.*, 2017).

1.4.3 CRISPR in the clinic

While *in vivo* editing is becoming more routine and model organisms are providing proof-of-principle evidence that at a molecular level many genetic diseases can be fixed by correcting the mutated gene, attention is turning to gene editing in the clinic and editing of human embryos (Plaza Reyes & Lanner, 2017). While editing human embryos is controversial, some institutes have been granted permission to modify human embryo's in the laboratory. The first account of editing human embryos was published in 2015 when Liang and colleagues edited tripronuclear zygotes; polyspermic zygotes that are generated during *in vitro* fertilisation and usually discarded (Liang *et al.*, 2015a). They reported editing efficiencies of ~52% in embryos that received Cas9 + sgRNA RNA molecules and 14% HDR editing when a ssODN molecule was co-injected. These low levels of editing, mosaic editing and observed off-target editing underscore the challenges that clinical gene editing faces and highlights the improbability of success using the current systems. A year later another group reported NHEJ rates of 47-63%, and HDR rates of 5% when using a ssODN or 7% when using a 1 kb plasmid (Kang *et al.*, 2016). Interestingly the 32 bp deletion that they were introducing by HDR was more efficiently inserted when two sgRNAs surrounding the region to be deleted were used, producing an editing efficiency of 15%. Again the resulting 6-18 cell embryos that were analysed were mosaic for the mutation, however no off-target editing was observed this time (Kang *et al.*, 2016).

Despite these limitations, CRISPR is already being used in clinical trials, with explanted cells being edited and returned to patients with the hope of alleviating disease symptoms. This was performed first by a group at Sichuan University in China who removed T cells from a patient with non-small cell lung cancer, used CRISPR to inactivate the *PD-1* gene, and then injected the edited T cells back into the patient (Sheridan, 2017). This approach is reminiscent of previous clinical trials using ZFNs to edit the *CCR5* gene in T cells from HIV patients (Tebas *et al.*, 2014). The ZFN introduced a 32bp deletion into the *CCR5* gene thus was the same deletion as was being tested in human tripronuclear zygotes described above (Kang *et al.*, 2016). While the edited T cells were insufficient in number to prevent further HIV infection, this clinical trial did demonstrate that ZFN edited T cells were safe for use in patients (Tebas *et al.*, 2014). In addition to the current CRISPR PD-1 trial, seven other programs are awaiting approval, targeting a wide range of diseases from sickle cell anaemia, multiple myeloma and beta-thalassemia. Most of these trials aim to perform the editing *ex vivo*, although some wish to tackle the more difficult challenge of delivering CRISPR components *in vivo* (Sheridan, 2017). While drug efficacy is also the aim for *in vivo*

targeting, safety is more of a concern due to the presence of a bacterial Cas9 protein being expressed in the human body with an unknown half-life.

Editing in human embryos has flagged up a number of concerns for *in vivo* editing, but many groups are pushing towards the use of CRISPR gene editing for treatment of genetic diseases. In this study, I will use CRISPR gene editing to target the *MECP2* gene in LUHMES neuronal progenitors, primarily for the purpose of creating model neuronal cell lines for study in the laboratory.

Project Aims

Project Aims

The major goal of this project is to determine if the LUHMES cell line and its derived human neurons will be a useful model system for studying MeCP2 and Rett syndrome. This goal can be broken down into a few main aims:

1. Characterise the LUHMES cell line in terms of differentiation dynamics and assess MeCP2 expression dynamics
2. Determine if the LUHMES cell line can be transfected, cloned and genetically manipulated using CRISPR technologies
3. Create a panel of MeCP2-mutated LUHMES cell lines
4. Analyse the morphological consequences of MeCP2 mutation and knock-out on LUHMES-derived neurons
5. Analyse gene expression changes in MeCP2 mutant LUHMES-derived neurons
6. Propose a mechanism for MeCP2 function with respect to transcriptional regulation
7. Assess the potential of protein transduction as a therapeutic strategy for RTT

Chapter 2

2. Materials and Methods

2.1 Buffers and Solutions

L Broth:

1% (w/v) Bacto tryptone, 0.5% (w/v) Bacto yeast extract, 1% (w/v) NaCl (Fisher Scientific 5/3160/63) (adjusted to pH 7.0 with 5 M NaOH)

Ampicillin:

Stock 50 mg/ml in water (Fisher Scientific T-22943)

Kanamycin:

Stock 20 mg/ml in water (Gibco 11815-024)

TE buffer:

10 mM Tris HCl (from 1M pH 7.5 stock; Invitrogen 15504-020), 1mM EDTA (VMR 20302.236; from 0.5M pH 8.0 stock)

1X TAE:

40 mM Tris (Invitrogen 15504-020), 1.142% (v/v) glacial acetic acid (Fisher Scientific A/0400/PB17), 1 mM EDTA (VMR 20302.236; from 0.5M pH 8.0 stock)

1X SDS running buffer:

24.8 mM Tris (Invitrogen 15504-020), 192 mM glycine (Sigma G8790), 1% (w/v) SDS (Sigma L4509)

1X Transfer buffer:

25 mM Tris (Invitrogen 15504-020), 190 mM glycine (Sigma G8790)

TBS:

Dissolve 0.5 M Tris (Invitrogen 15504-020) and pH to 8.0 with HCl (Fisher Scientific H/1200/PB17), add 1.5 M NaCl (Fisher Scientific 5/3160/63).

PBS:

1 tablet (Fisher BR14a) per 100ml

HBS:

16g NaCl, 0.74g KCl, 0.252g Na_2PO_4 (dihydrate), 2g D-Glucose (Dextrose), 10g HEPES with 180ml H_2O, pH to 7.2, top up with H_2O to 200ml and filter-sterilised. 1X HBS made using sterile H_2O.

Ponceau S:

0.1% (w/v) ponceau (Sigma P 7767) in 5% acetic acid (Fisher Scientific A/0400/PB17) in water.

HBSS + HEPES buffer for calcium imaging:

Remove 10 ml from a 500 ml bottle of HBSS (either with or without calcium + magnesium). Add 10 ml of 1 M HEPES (Sigma, 3537).

Calcium buffers for calcium imaging:

0 mM Ca^{2+} solution: 5.2 ml of 5M NaCl, 360 µl of 3M KCl, 100 µl of 1M $MgCl_2$, 0.076g EGTA + ddH_2O to 200ml.

8mM Ca^{2+} solution: To 50 ml of 0 mM solution add 400 µl of 1M $CaCl_2$.

For range in between serial dilute the 0 mM and 8 mM Ca^{2+} solutions to create 6 mM, 4 mM, 2 mM, 1 mM, 0.5 mM, 0.25 mM and 0.125 mM solutions.

EGTA for negative control for calcium imaging:

Make a 40 X dilution by dissolving 0.936 g EGTA powder in 5 ml distilled water. Add 750 µl of 40X EGTA to 15 ml HBSS+Ca^{2+}+Mg^{2+}+HEPES to make 2X EGTA. Aliquot into 1 ml Eppendorfs and store at -20^{o}C. Do not repeatedly freeze-thaw.

A23187 Calcium ionophore for positive control for calcium imaging:

To the 1 mg stock add 710 µl of DMSO. Take 20 µl of the reconstituted ionophore and add it to 2 ml HBSS+Ca^{2+}+Mg^{2+}+HEPES to make a 2X calcium ionophore stock. Aliquot as 50 µl and store at -20°C.

KCl for calcium imaging:

0.29 g in 30 ml HBSS+Ca^{2+}+Mg^{2+}+HEPES gives a 130 mM KCl stock.

0.58 g in 30 ml HBSS+Ca^{2+}+Mg^{2+}+HEPES gives a 260 mM KCl stock.

2.2 Primary antibodies

Antibody	**Company + Catalogue Number**	**Experiment**	**Dilution**	**Species**	**RRID**
MeCP2	Sigma 7443	WB; IF	1000X	Mouse monoclonal	AB_477235
MeCP2	Sigma 6818	WB	1000X	Mouse monoclonal	AB_262075
MeCP2	Cell signalling D4F3	WB; IF	1000X	Rabbit monoclonal	AB_2143849
GAPDH	Cell Signalling 5174	WB	5000X	Rabbit monoclonal	AB_10622025
H3	Abcam 1791	WB	5000X	Rabbit polyclonal	AB_302613
mCherry	Abcam 167453	WB, IF	1000X	Rabbit polyclonal	AB_2571870
Neurofilament	Covance SMI-311R-100	IF	500X	Mouse monoclonal	AB_509991
MAP2	Abcam 5392	IF	5000X	Chicken polyclonal	AB_2138153
NeuN	Millipore, ABN78	IF	500X, 1000X	Rabbit polyclonal	AB_10807945
Tyrosine Hydroxylase	Millipore, AB152	IF	200X	Rabbit polyclonal	AB_390204
DAT	Millipore, MAB369	IF	1000X	Rabbit monoclonal	AB_2190413
LaminB1	Abcam, ab16048	IF	500X	Rabbit polyclonal	AB_443298
HDAC3	Sigma, 3E11	WB	1000X	Mouse monoclonal	AB_1841895
TBLR1	Abcam, ab24548	WB	1000X	Rabbit polyclonal	AB_2199904

Table 1. Details of primary antibodies used in this study.

All primary antibodies are listed along with company and catalogue information, the type of experiment they were used in, the final concentration and RRID. RRID = research resource identifier (https://scicrunch.org/resources). WB = western blot. IF = immunofluorescence.

2.3 Primer sequences

All primers were ordered as standard desalted oligonucleotides from Sigma and were reconstituted in Nuclease-free H_2O (Ambion, AM9937) to make 100 µM stocks that were stored at -20°C. Working dilutions of 10 µM or 3.2 pM were stored at 4°C.

Primer name	**Sequence**	**Experiment**
hMeCP2_1_F	GATCAATCCCCAGGGAAAAG	Various PCR + sequencing
hMeCP2_1_R	CCTCTCCCAGTTACCGTGAA	Various PCR + sequencing
hMeCP2_2_F	GAGACCGTACTCCCCATCAA	Various PCR + sequencing
hMeCP2_2_R	AGTCCTTTCCCGCTCTTCTC	Various PCR + sequencing
hMeCP2_3_F	CAAGGCCAAACAGAGAGGAG	Various PCR + sequencing
hMeCP2_3_R	CAATCCGCTCCGTGTAAAGT	Various PCR + sequencing
hMeCP2_4_F	AGCCAAATGACTGAAAGCACT	Various PCR + sequencing
hMeCP2_4_R	CCATACGGGTCTGGAGCAG	Various PCR + sequencing
hMeCP2_5_F	AGCTCCTTGTCAAGATGCCT	Various PCR + sequencing
hMeCP2_5_R	CCAACTACTCCCACCCTGAA	Various PCR + sequencing
hMeCP2_6_F	CGCTCTGCTGGGAAGTATGA	Various PCR + sequencing
hMeCP2_6_R	GCTACTTCTGGCCCTGGTTA	Various PCR + sequencing
hMeCP2_7_F	AGGGCTCAGGGAAGAAAAGT	Various PCR + sequencing
hMeCP2_7_R	CTTTTCCCTGGGGATTGAT	Various PCR + sequencing
hMeCP2_8_F	GGGCTCAGGGAAGAAAAGTC	Various PCR + sequencing
hMeCP2_8_R	GCAGAAATGGAAGGGGAGAA	Various PCR + sequencing
hMeCP2_9_F	TGTTAGGGCTCAGGGAAGAA	Various PCR + sequencing
hMeCP2_9_R	GGACGGAGGAAGGGAAAGAA	mCherry PCR screens
hMeCP2_exon1_F	GGAGAGAGGGCTGTGGTAA	Various PCR + sequencing
hMeCP2_exon1_R	AGTCTCTCCTCCTCGCCTC	Various PCR + sequencing
hMeCP2_exon2_F	GCTCCATAAAAATACAGACTCA	Various PCR + sequencing
hMeCP2_exon2_R	CTGAGCCCTAACATCCCAG	Various PCR + sequencing
hMeCP2_exon3_F	GGAAGAAAAGTCAGAAGACCA	Various PCR + sequencing
hMeCP2_exon3_R	TTGATCAAATACACATCATACTT	Various PCR + sequencing
hMeCP2_exon4_F	TCCCCAGGGAAAAGCCTTTC	Various PCR + sequencing
hMeCP2_exon4_R	AGCTGCCTTTATTCTTGTTGG	Various PCR + sequencing
hMeCP2_intron2_F	CCACAGCCCAAATTCCTAAA	Various PCR + sequencing
hMeCP2_intron2_2F	TCCCTTGAAGTGCGACTCAT	mCherry PCR screens
hMeCP2_intron3_F	GTCTTTCTGTTTGTCCCCACA	Various PCR + sequencing
hMeCP2_intron3_2F	CAGACGAGTGAGTGGCTTTG	Various PCR + sequencing
V1_full_F	ATGGTAGCTGGGATGTTAGG	Various PCR + sequencing
V1_full_R	TCAGCTAACTCTCTCGGTCA	Various PCR + sequencing
hGAPDH_3_F	CAGCCTCAAGATCATCAGCA	qPCR

hGAPDH_3_R	TGTGGTCATGAGTCCTTCCA	qPCR
pLKO1_5	GACTATCATATGCTTACCGT	pCRISPR sequencing primer
SNAPC4_F	CACAGTACATGGTCCCGGAC	Off-target locus for sgRNA16
SNAPC4_R	CTTTGTGGCTGGCACTCTTT	Off-target locus for sgRNA16
SNAPC4_F2	CCAATGACTCCCAGGACAGA	Off-target locus for sgRNA16
MYOZ1_F	AACTGTCGCTGCTTACCAAC	Off-target locus for sgRNA16
MYOZ1_R	TCGGCTTTCATCTGCTCTCA	Off-target locus for sgRNA16
MYOZ1_F2	ACAGCTGGTCAGGGATTCTC	Off-target locus for sgRNA16
MYOZ1_R2	ACCACCCCTACCTGTTGAAG	Off-target locus for sgRNA16
KIAA1324_F	ACAGAGGGAGGAGGAAGGAT	Off-target locus for sgRNA16
KIAA1324_R2	TCATCCCTCTCGCACATCTC	Off-target locus for sgRNA16
KIAA1324_R	GGCAATGTTTCTCACCAGCA	Off-target locus for sgRNA16
MAP7D3_F	CACGGAACATAGGAGAGGGT	Off-target locus for sgRNA17
MAP7D3_F2	GCAGCCCATCCTTGAGAGA	Off-target locus for sgRNA17
MAP7D3_F3	CGGCAGCCCATCCTTGAG	Off-target locus for sgRNA17
MAP7D3_R	TCTGAAGTGCTGATTCCCGT	Off-target locus for sgRNA17
MAP7D3_R2	GAGGACCAGATAGCACGAGG	Off-target locus for sgRNA17
MAP7D3_R3	CCAGGGAAAAGTTTTGTGGC	Off-target locus for sgRNA17
TRPV3_F	TGTGAGTGGCATCTGGTGAT	Off-target locus for sgRNA17
TRPV3_R	AACCCCGCGTATAGTAGAGC	Off-target locus for sgRNA17
ZCCH14_F	TTTAGCACTGGAAGCCTCAG	Off-target locus for sgRNA15
ZCCH14_R	GGAGTTGAACCGTTGACACC	Off-target locus for sgRNA15
WNK1_F	GGTTTCTCAGCAGCCGTTAA	Off-target locus for sgRNA15
WNK1_R	AGCCGACCGTTTCTATTCCT	Off-target locus for sgRNA15
MMP14_F	ACAGCTATCCTTTGCCCACT	Off-target locus for sgRNA15
MMP14_R	GTTTTCCACACAGCACTCCC	Off-target locus for sgRNA15
hSox2_f	CAAGATGCACAACTCGGAGA	qPCR
hSox2_r	GCTTAGCCTCGTCGATGAAC	qPCR
hNestin_f	TGCGGGCTACTGAAAAGTTC	qPCR
hNestin_r	GAGCGATCTGGCTCTGTAGG	qPCR
Ki67_2_F	TGGGCGAAGTTCACAGTCAA	qPCR
Ki67_2_R	TGAGCACTCTGTAGGGTCGA	qPCR
hCDK6_f	TGCACAGTGTCACGAACAGA	qPCR
hCDK6_r	ACCTCGGAGAAGCTGAAACA	qPCR
PAX6_f	AGCCCCATATTCGAGCCCCG	qPCR
PAX6_r	TCGGGAAATGTCGCACGGCC	qPCR
hPax3_f	AGGCATGGATTTTCCAGCTA	qPCR
hPax3_r	AGTCTGGGGCTGATGAGGTA	qPCR
hNEUROD1_1_f	GCCCCAGGGTTATGAGACTA	qPCR
hNEUROD1_1_r	CTCGCTGTACGATTTGGTCA	qPCR
hNDNF_f	CAGATGGAGCTGAAATTAGCA	qPCR

hNDNF_r	CTCAGCTTCCACTCCAAAGG	qPCR
TUBB3_f	CGCCCAGTATGAGGGAGAT	qPCR
TUBB3_r	AGTCGCCCACGTAGTTGC	qPCR
hFox-3_f	CCGACCCTACAGAGAAGCAG	qPCR
hFox-3_r	GAATTGCCCGAACATTTGC	qPCR
hDCX_1_f	ACCGCTACTTCAAGGGGATT	qPCR
hDCX_1_r	AGACAGAGATCGCGTCAGGT	qPCR
hSEZL6_1_f	TGTACTCCCACCCCTACAGC	qPCR
hSEZL6_1_r	TTCCCCTGTCTCGTAAATGG	qPCR
hSEZL6_2_f	GATCCAGAATGGCTGGAAAA	qPCR
hSEZL6_2_r	ACTCCCCACGATGTCATAGC	qPCR
hSEMA4D_f	TGCAGTGATGTTTGGGACAG	qPCR
hSEMA4D_r	TGTCTGGCTCATGAAACTGC	qPCR
SYP_f	TGCCAACAAGACCGAGAGTG	qPCR
SYP_r	TTCGGCTGACGAGGAGTAGT	qPCR
SYN1_f	TCAGACCTTCTACCCCAATCA	qPCR
SYN1_r	GTCCTGGAAGTCATGCTGGT	qPCR
DLG4_f	CACTCCTCACAGTGCTGCAT	qPCR
DLG4_r	TGTCTTCATCTTGGTAGCGG	qPCR
GRIN1_F	AAGCCCAACGCCATCCAGAT	qPCR
GRIN1_R	GAGTGAAGTGGTCGTTGGGGGTA	qPCR
NLGN1_f	ATGATGGAAGTGTCTTGGCA	qPCR
NLGN1_r	ATAGTTCCCCTTTGCAGCCT	qPCR
SV2A_F	AGATCGGCAGGCTCAGAAT	qPCR
SV2A_R	GACAGGAAGAAGCAGGAGACA	qPCR
SNAP25_f	CTGTCTTTCCTTCCCTCCCT	qPCR
SNAP25_r	GGGTCAGTGACGGGTTTG	qPCR
hTH_f	GTGTTCCAGTGCACCCAGTA	qPCR
hTH_r	GCCAATGTCCTGCGAGAA	qPCR
hDAT_f	AGTGGCCTGGTTCTATGGTG	qPCR
hDAT_r	GACCACGAACAGGAGAAAGC	qPCR
hDRD2_f	GGAGGTGGTAGGTGAGTGGA	qPCR
hDRD2_r	GATGCTGATGGCACACAAGT	qPCR
hEN1_f	AAAACTGACTCGCAGCAGCC	qPCR
hEN1_r	TGCTCCGTGATGTAGCGGTT	qPCR
hSHH_f	CCAATTACAACCCCGACATC	qPCR
hSHH_r	CAGTTTCACTCCTGGCCACT	qPCR
VMAT-2_f	TGGGGAGGTGGCTTTGTGCT	qPCR
VMAT-2_r	CCCATAGACGGACACGTGCC	qPCR
hGIRK2_F	GACCTGCCAAGACACATCAG	qPCR
hGIRK2_R	CGGTCAGGTAGCGATAGGTC	qPCR

Cre_1_F	GGCGTTTTCTGAGCATACCT	AAV testing
Cre_1_R	ATCTTCAGGTTCTGCGGGAA	AAV testing
Cre_2_F	ATTGGCAGAACGAAAACGCT	AAV testing
Cre_2_R	ATCAGCTACACCAGAGACGG	AAV testing
EGFP_1_F	GGTGAACTTCAAGATCCGCC	AAV testing
EGFP_1_R	GGTGCTCAGGTAGTGGTTGT	AAV testing
EGFP_2_F	CAAAGACCCCAACGAGAAGC	AAV testing
EGFP_2_R	TCCATGCCGAGAGTGATCC	AAV testing
attB1_hPGK	GGGGACAAGTTTGTACAAAAAAGCAGGC TTTGGGGTTGGGGTTGCGCCTT	Gateway cloning for pAAV PGK-puro plasmid
attB4_hPGK	GGGGACAACTTTGTATAGAAAAGTTGGGT GCTGGGGAGAGAGGTCGGTG	Gateway cloning for pAAV PGK-puro plasmid
attB4r_Puro	GGGGACAACTTTTCTATACAAAGTTGTTA TGACCGAGTACAAGCCCAC	Gateway cloning for pAAV PGK-puro plasmid
attB3r_Puro	GGGGACAACTTTATTATACAAAGTTGTTC AGGCACCGGGCTTGCG	Gateway cloning for pAAV PGK-puro plasmid
attB1_MeCP2_intron2F	GGGGACAAGTTTGTACAAAAAAGCAGGC TCCACAGCCCAAATTCCTAAA	Gateway cloning for pAAV mCherry targeting plasmid
attB4_MeCP2_stopR	GGGGACAACTTTGTATAGAAAAGTTGGGT GGCTAACTCTCTCGGTCACGG	Gateway cloning for pAAV mCherry targeting plasmid
attB3_MeCP2_stopF	GGGGACAACTTTGTATAATAAAGTTGTGA CTTTACACGGAGCGGAT	Gateway cloning for pAAV mCherry targeting plasmid
attB2_MeCP2_8R	GGGGACCACTTTGTACAAGAAAGCTGGGT AGCAGAAATGGAAGGGGAGAA	Gateway cloning for pAAV mCherry targeting plasmid
mCherry_IF_Tm_B4r	GGGGACAACTTTTCTATACAAAGTTGTTA TGGTGAGCAAGGGCGA	Gateway cloning for pAAV mCherry targeting plasmid
mCherry_IF_Tm_B3r	GGGGACAACTTTATTATACAAAGTTGTCT ACTTGTACAGCTCGTCCATGC	Gateway cloning for pAAV mCherry targeting plasmid
pAAV_Gateway_F	AAGCAGCGTATCCACATAGC	Check final pAAV sequence
pAAV_Gateway_F2	TGATTGCCCCACCATTTTGT	Check final pAAV sequence
pAAV_Gateway_F3	AGAGTTCTTGCAGCTCGGTGAC	Check final pAAV sequence
pAAV_Gateway_R	CTTCTCCTCCGGGCTGTAAT	Check final pAAV sequence
pAAV_Gateway_R2	GAATCACCGACCTCTCTCCC	Check final pAAV sequence
pAAV_Gateway_R3	CAACCTCCCCTTCTACGAGC	Check final pAAV sequence
R306C_F	TCCCCATCAAGAAGTGCAAGACCCGGGA G	Site-directed mutagenesis of R306C targeting plasmid
R306C_R	CTCCCGGGTCTTGCACTTCTTGATGGGGA	Site-directed mutagenesis of R306C targeting plasmid
5_6_PAM_F	ATCGAGGTCAAAGAAGTGGTGAAGCCCCT GCTAGTGTCCACCCTCG	Site-directed mutagenesis of R306C targeting plasmid
5_6_PAM_R	CGAGGGTGGACACTAGCAGGGGCTTCACC ACTTCTTTGACCTCGAT	Site-directed mutagenesis of R306C targeting plasmid

R111G_F	GACACGGAAGCTTAAGCAAGGGAAATCT GGCCGC	Site-directed mutagenesis of R111G plenti
R111G_R	GCGGCCAGATTTCCCTTGCTTAAGCTTCC GTGTC	Site-directed mutagenesis of R111G plenti
mCherry_BamH1_F	ACCGAGAGAGTTAGCTGTAAGG	Clone mCherry into pRRLSIN
mCherry_Sal1_R	GTGCGTCGACTTATACTTGTACAG	Clone mCherry into pRRLSIN

Table 2. Sequences of primers used in this study.

All primers used in this study are listed along with their sequence and the application they were used for.

2.4 Mammalian Tissue Culture

All tissue culture was performed at 37°C in a humidified 5% CO_2 atmosphere. LUHMES tissue culture medium and methods were as described in Scholz *et al*, 2011, with some minor alterations. LUHMES cells were propagated in Nunclon cell culture flasks (Thermo Scientific) that were pre-coated with 43 µg/ml poly-L-ornithine (Sigma) and 1 µg/ml fibronectin (Sigma) in H_2O overnight at 37°C. The next day coating solution was aspirated; flasks were washed with H_2O and air-dried before cell seeding. Vessels for differentiation were used fresh, however vessels for proliferating LUHMES cells could be stored post-coating at 4°C for up to one week. Cells were routinely seeded at a density of $2x10^6$ cells/T75, and passaged every 2 days. LUHMES cell proliferation medium consists of Advanced DMEM/F12 (Gibco, 12634) supplemented with 2 mM L-Glutamine (Sigma), 1X N2 (Gibco) and 40 ng/ml β-FGF (R&D Systems) while differentiation medium consisted of Advanced DMEM/F12 supplemented with 2 mM L-Glutamine, 1X N2, 1 mM cAMP (Sigma), 1 µg/ml tetracycline (Sigma) and 2 ng/ml GDNF (R&D Systems). All proliferation and differentiation medium was filtered through a 0.22 µm PES filter (Millex GP, SLGP033RS) prior to use and could be stored for up to one week. When culturing the puromycin-resistant, lentivirus-infected LUHMES cell lines, 0.25 µg/ml puromycin was included in the proliferation medium, however no puromycin was added to differentiation medium. For cell detachment, cells were incubated with Trypsin (Invitrogen) for 1 minute at 37°C, collected in Advanced DMEM/F12 medium and centrifuged at 1300 rpm for 5 minutes at room temperature. The cell pellet was resuspended in Advanced DMEM/F12 medium, counted using a Scepter device (Millipore) and seeded at the appropriate density (Table 3). When seeding cells on 19 mm coverslips (VWR, Thickness No 1.5) for immunofluorescent imaging coverslips were placed in 12-well plates and coated in 1 ml of coating solution; day 2 neurons were seeded as 0.15 x 106 cells in a 150 µl drop onto the coverslip itself. Cells were left in the incubator for 10 minutes to allow cell attachment to the coverslip, then 1 ml

of differentiation media was added to each well. The same approach was used for seeding 1 x 106 day 2 cells on 25 mm coverslips (Harvard Apparatus UK, Thickness No 1.5) in 6-well plates for calcium imaging. During all differentiations a half-media change was performed on day 6 and every 3 days after this.

Plastic wear	Surface area (cm^2)	Cells for 2 days Pro	Cells for 3 days Pro	Cells for 4 days Pro	Cells for Diff media change	Cells for Diff seed	Coating volume	Medium volume	Trypsin volume	Post trypsin resuspension volume
96-well plate	0.36						100ul	100/200ul	50ul	
24-well plate	1.8					$0.2x10^6$	500ul	500ul	75ul	1ml
12-well plate	3.5	$0.1x10^6$				$0.35x10^6$ (0.15 on a coverslip)	1ml	1ml	150ul	1ml
6-well plate	9.6	$0.25x10^6$				$1x10^6$	2ml	2ml	200ul	2ml
T75 flask	75	$2x10^6$	$1x10^6$		$2.5x10^6$	$8x10^6$	8ml	10ml	4ml	10ml
T175 flask	175	$4.8x10^6$			$6x10^6$	$20x10^6$	16ml	30ml		20ml
6cm dish	21					$2x10^6$	4ml	4ml		
10cm dish	60	$1.6x10^6$	$0.8x10^6$	$0.35x10^6$	$2x10^6$	$6x10^6$	6ml	8ml	3ml	8ml
15cm dish	140	$3.7x10^6$				$16x10^6$	12ml	25ml		20ml

Table 3. Volumes and cell numbers for LUHMES tissue culture.

Hek293FT cell culture medium consisted of DMEM (Gibco, 49166) medium supplemented with 10% fetal bovine serum (Gibco), 2 mM L-Glutamine (Gibco), 10X non-essential amino acids (Gibco), 100 U/ml penicillin (Gibco) and 100 µg/ml streptomycin (Gibco). Cells were propagated in T75 cell culture flasks in 20ml complete medium, and were split 1:10 or 1:20 when reaching 80-90% confluency by washing with PBS, trypsinisation and centrifugation.

2.5 Plasmids and cloning: CRISPR, lentiviral, AAV

All CRISPR plasmids were purchased from Addgene and originated from the Feng Zhang lab (pX330 #42230, pX335 #42335, pX458 #48138, plentiCRISPR #52961). The crispr.mit.edu online webtool was used for sgRNA design (Table 4). Cloning of sgRNAs

into CRISPR plasmids was performed following the protocol from the Zhang lab, available online at genome-engineering.org and described in ref (Cong *et al.*, 2013). All ssODNs were ordered from Sigma as desalted oligonucleotides (Table 5).

Gateway cloning (Invitrogen) was used to create all AAV plasmids, as well as the mCherry targeting plasmid and the pPuro puromycin resistance plasmid. The pmaxGFP plasmid was provided with the Nucleofection kit bought from Lonza. Lentiviral packaging plasmids were bought from Addgene from Didier Trono; pMD2.G (Addgene #12259) and psPax2 (Addgene #12260). The transfer plasmids for the mutant MeCP2 over-expression viruses were based on a pLKO.1 backbone (from Bob Weinberg Addgene #84530) and were cloned by performing site-directed mutagenesis using the pLKO_CMV_MeCP2v2_WPRE plasmid created by Dr Justyna Cholewa-Waclaw. The transfer plasmid for the mCherry expressing lentivirus was created by amplifying a BamH1-mCherry-Sal1 construct from pBSMe2NSmCherry (gift from Dr Jacky Guy), and ligating it into pRRLSIN.cPPT.PGK-GFP (gift from Dr Michael Robson).

The R306C targeting plasmid was created by firstly performing a PCR reaction using Phusion polymerase in GC buffer (NEB) with approximately 100-250 ng LUHMES genomic DNA in a 25 µl reaction volume. Subcloning was performed using 2 µl of this PCR mix with the Zero Blunt TOPO PCR cloning kit (Invitrogen) and 2 µl of the subcloning reaction was transformed into DH5α cells. The following day two colonies were picked into 3 ml cultures of LB/kanamycin and grown overnight by shaking at 37ºC. The next day plasmids were extracted from 1.5 ml of culture using the Plasmid Miniprep kit (Qiagen) and sequenced using primers spanning the *MECP2* locus. After sequencing confirmed error-free incorporation of the *MECP2* fragment into the pTOPO vector, 50 ml cultures of LB/kanamycin were set up using the remaining 1.5 ml of bacterial culture and grown overnight by shaking at 37ºC. The next day plasmids were extracted using the Plasmid Maxiprep kit (Qiagen). This plasmid was subjected to site-directed mutagenesis using the QuickChange II XL kit (Agilent) following the manufacturer's instructions. Two rounds of mutagenesis PCRs were used to incorporate first the R306C point mutation, and then two PAM-abolishing point mutations for two separate sgRNAs to create the final targeting vector.

sgRNA name	Experiment	Sequence (5' → 3')
sgRNA 1	Knock-in of *R306C-MECP2*	GTCTTCTATCCGATCTGTGC
sgRNA 2	Knock-in of *R306C-MECP2*	GTACGGTCTCCTGCACAGAT
sgRNA 3	Knock-in of *R306C-MECP2*	GCGGGTCTTGCGCTTCTTGAT
sgRNA 4	Knock-in of *R306C-MECP2*	GCCGGGAGACGGTCAGCATCG
sgRNA 5 (2)	Knock-in of *R306C-MECP2*	GACGGTCAGCATCGAGGTCA
sgRNA 6	Knock-in of *R306C-MECP2*	GGAAGTGGTGAAGCCCCTGC
sgRNA 11 (C)	Knock in of mCherry tag into *MECP2*	GTTAGCTGACTTTACACGGAG
sgRNA 12	Knock in of mCherry tag into *MECP2*	GCAGCTAACTCTCTCGGTCAC
sgRNA 13	Knock in of mCherry tag into *MECP2*	GACGGGCGTCCGGCTGTCCAC
sgRNA 15 (1)	Knock-in of *R306C-MECP2*	GCCATCAAGAAGCGCAAGACC
sgRNA 16 (A)	Knock-out of MeCP2	GAGAAGCTTCCGGCACAGCCG
sgRNA 17 (B)	Knock-out of MeCP2	GCGCTCCATCATCCGTGACCG
sgRNA K1	Knock-in of *R111G-MECP2*	GAGGAAATCTGGCCGCTCTGC
sgRNA K2 (3)	Knock-in of *R111G-MECP2*	GGACACGGAAGCTTAAGCAA
sgRNA K3	Knock-in of *R133C-MECP2*	GTTCGCTCTAAAGTGGAGTTG
sgRNA K4 (4)	Knock-in of *R133C-MECP2*	GAAAAGCCTTTCGCTCTAAAG
sgRNA K5 (5)	Knock-in of *T158M-MECP2*	GATTTTGACTTCACGGTAAC
sgRNA K6 (6)	Knock-in of *T158M-MECP2*	GATTTTGACTTCACGGTAACT

Table 4. Sequences of sgRNAs used in this study.

All sgRNAs have 20 nucleotides of homology to the *MECP2* locus, however some sgRNAs have an extra G nucleotide added at the 5' terminus for efficient U6 promoter transcription and are therefore 21 nucleotides long. Numbers and letters in brackets refers to the names given to these sgRNAs in Chapter 4.3.

ssODN	Experiment	Sequence (5' → 3')
1	Knock-in of *R306C-MECP2*	GAGTCTTCTAT**T**CGATCTGTGCA**A**GAGACCGTACTCCCCATCAAGAAG**T**GCAAGACCC GGGAGACGGTCAGCATCGA**A**GTCAA**A**GAAGTGGTGAAGCCCCTGCT**A**GTGTC
2	Knock-in of *R306C-MECP2*	GACACCAGCAGGGGCTTCACCACTTC**T**TTGACCTCGATGCTGACCGTCTCCCGGGTCTT GC**A**CTTCTTGATGGGGAGTACGGTCTCCTGCACAGATCGGAT**C**GAAGACTC
3	Knock-in of *R306C-MECP2*	GACACCAGCAGGGGCTTCACCACTTCCTTGACCTCGATGCTGACCGTCTC**G**CGGGTCT TGC**A**CTTCTTGATGGGGAGTACGGTCTCCTGCACAGATCG**A**ATAGAAGACTC
4	Knock-in of *R111G-MECP2*	CTTACTTACTTGATCAAATACACATCATACTTCCCAGCAGAGCG**T**CCAGATTT**G**C**C**TTG CTTAAGCTTCCGTGTCCAGCCTTCAGGCAGGGTGGGGTCAT
5	Knock-in of *R133C-MECP2*	CAGGGATGTGTCGCCTACCTTTTCGAAGTACGCAATCAA**T**TC**G**ACTTTAGAGC**A**AAAT GCTTTTCCCTGGGGACTGTGGGGACAAACAGAAAGACACAAG
6	Knock-in of *T158M-MECP2*	GGCTTCTTAGGTGGTTTCTGCTCTCGCCGGGAGGGGCTCCCTCT**G**CCAGTTACC**A**TGA AGTCAAAATCATTAGGGTCCAGGGATGTGTCGCCTACCTTTT

Table 5. Sequences of ssODNs used in this study.

Nucleotides highlighted in bold differ from the WT *MECP2* sequence, and the single bold red nucleotide in each ssODN denotes the point mutation of interest. Underlined 5'-CGA-3' sequence

in ssODN 2 codes for a Serine residue, whose WT sequence would be 5'-TGA-3'. The T→C base change is an error in the sequence.

2.6 Cell transfections

Hek293FT cells were transfected using Lipofectamine-2000, the amount of which depended on the number of cells to be transfected (Table 2.6). The protocol consisted of mixing optiMEM solution (Gibco) with Lipofectamine-2000 (Thermo Fisher) and letting this sit at room temperature for 5 minutes. Then a DNA:optiMEM mix was added dropwise into the Lipofectamine-2000:optiMEM mix and this was left for at least 20 minutes at room temperature. Meanwhile Hek293FT cells were trypsinised and counted using a Scepter device (Millipore). The appropriate number of cells was seeded using medium absent of penicillin and streptomycin. Immediately after seeding cells (i.e. while cells are still in suspension) the DNA:Lipofectamine-2000:optiMEM mix was added dropwise onto the cells and mixed by swirling. The following day medium was changed for fresh medium containing penicillin and streptomycin.

Culture vessel	**6-well plate**	**10cm dish**
Number of Hek293FT	1 x 10^6	6 x 10^6
OptiMEM volume	150 µl	1.5 ml
Lipofectamine-2000 volume	7.5 µl	36 µl
Approximate DNA amount	2.5 µg	12 µg

Table 6. Lipofectamine-2000 transfection of Hek293FT cells.

Test transfections of LUHMES cells were performed using a Neon Transfection system (Invitrogen) following the manufacturer's instructions. Briefly, cells were trypsinised and washed in PBS prior to mixing with plasmid DNA. Cells and DNA were then transferred into the Neon pipette, transfected via an electrical pulse and immediately seeded into a 24-well plate.

Test transfections of LUHMES cells using JetPrime (Polyplus) were performed according to the manufacturer's instructions.

Test electroporation's of LUHMES cells were performed using a GenePulser XCell electroporation machine (Bio-Rad). Each electroporation contained 1 µg of pmaxGFP (Lonza), 1 x 10^6 cells and 600 µl of ice-cold 1X HBS, pH 7. Immediately after electroporation cells were rested at room temperature for 2 minutes before seeding into a well of a 6-well plate.

LUHMES cells were routinely transfected by Nucleofection (Lonza) using a Basic Nucleofector kit for primary neurons (Lonza, VAPI-1003) and a Nucleofector II device. LUHMES cells were dissociated with 4 ml of trypsin, centrifuged at 13000 rpm for 5 minutes and resuspended in PBS for cell counting using a Scepter device. Aliquots of $2x10^6$ cells were pipetted into 15ml Falcons and these were centrifuged at 13000 rpm for 5 minutes. PBS was removed from all 15ml Falcons and the appropriate volume of each plasmid/ssODN was added to each tube. One by one cells and plasmids were then resuspended in 90µl of the Nucleofection solution and immediately transferred into a cuvette (provided in the kit) for electroporation with program D-33 unless otherwise stated in the results section. After electroporation, RPMI medium (Sigma) was added using the pipette provided in the kit and cells were moved into 15ml Falcons and incubated at 37^{o}C for 5 minutes before plating out into 6-well plates containing pre-warmed LUHMES proliferation media. Media was changed after a minimum of 4 hours.

2.7 FACS cytometry analysis of LUHMES cells

For flow cytometry sorting of single LUHMES cells into individual wells of a 96-well plate, 96-well plates (Greiner) were pre-coated overnight in PLO/fibronectin as described in section 2.1. Proliferation medium was supplemented with 1X B27 (Sigma), 100 U/ml penicillin (Gibco) and 100 µg/ml streptomycin (Gibco) and 100 µl was added to each well. Plates were stored in a plastic box kept in the 37^{o}C incubator while cells were harvested. LUHMES cells were trypsinised and centrifuged in a 15ml Falcon tube as described in section 2.1 and resuspended in 1 ml of Advanced DMEM/F12 supplemented with 10 µM HEPES (Sigma, H3537). Cells, 96-well plates and spare Advanced DMEM/F12 + HEPES were taken over to the Ashworth FACS facility and cells were sorted using a 100 µm nozzle with a FACSaria (BD Biosciences) machine at room temperature by Dr Martin Waterfall.

For analysis of fluorescence only, the LSRFortessa (BD Biosciences) machine was used with the help of Dr Martin Waterfall. Neurons were detached by treatment with ESGRO Complete Accutase (Millipore) for 3 minutes followed by centrifugation at 1300 rpm for 5 minutes at room temperature. Neurons were resuspended in Advanced DMEM/F12 for flow cytometry analysis. For assessment of dead cells, samples were incubated with LIVE/DEAD Fixation Far-Red Dead Cell Stain Kit (Life Technologies, L10120) at a concentration of 1:5000 for 30 minutes. The dye was washed out by centrifugation prior to flow cytometry analysis.

2.8 Targeting pipeline in LUHMES cells

Day 1 – Thaw low passage number LUHMES

Day 3 – Passage once

Day 5 – Nucleofect and change media after 4 hours

Day 7 – Take for FACS sorting of single GFP-positive cells into a 96-well plate

Day 13 – Top up 96-well plate with 100ul of media

Day 16 – Start moving clones from 96-well plates to 24-well plates

Day 17 onwards – Split individual clones in a 24-well plate as they become confluent, freeze half the well for liquid nitrogen, move the other half to a new well in a 24-well plate.

2.9 Adeno-Associated Virus production and infection

Adeno-associated viruses with serotype AAV1/2 were produced as described in Reference (McClure *et al.*, 2011). To produce rAAV-rh10 virus, pRV1 (rAAV1-specific) and pH21 (rAAV2-specific) were replaced with pAAV-rh10 plasmid (Penn Vector Core, School of Medicine, gene therapy program, University of Pennsylvania). To test AAV aliquots on Hek293FT or LUHMES cells, cells were grown in 6-well plates and viruses were thawed quickly in a 37^{o}C water-bath, vortexed briefly and the desired volume of virus was added to 2 ml aliquots of proliferation medium. Medium was aspirated from cells in the 6-well plate and replaced with medium containing virus. After 24 hours of incubation medium was again changed to normal proliferative medium.

2.10 Lentivirus production and infection

To produce lentiviruses Hek293FT cells were transfected with lipofectamine-2000 as described in section 2.3. Each transfection contained 6 x 10^6 cells, 4.6 µg of psPax2 2nd generation packaging plasmid, 2.8 µg of pMD2.G VSV-G envelope expressing plasmid and 7.5 µg of transfer plasmid. Media was changed 24 hours later and initial expression of fluorescent reporters (when present in the transfer plasmid) was determined by UV microscopy to confirm expression. A further 48 hours later cell media was transferred to 15 ml Falcon tubes and centrifuged at 3200 xg for 15 minutes at 4^{o}C. Supernatant was poured into new 15ml Falcon tubes and 2.5 ml PEG solution from the PEG virus precipitation kit (BioVision) was added and thoroughly mixed before storage at 4^{o}C overnight. The next day 15 ml Falcons were centrifuged at 3200 xg for 30 minutes at 4^{o}C. Pellets were resuspended in 200 µl virus resuspension solution (BioVision) and stored in 100 µl aliquots at -80^{o}C.

For test infections of lentiviruses that confer puromycin resistance, LUHMES cells were seeded at 0.2 x 10^6 cells/well in a 6-well plate two days prior to infection. On the day of infection (day 0) the vials of viruses to be tested were put on dry ice and thawed quickly in a 37^{o}C waterbath, straight from dry ice. Once the viruses were thawed, they could be kept at room temperature for as long as was necessary. Media was changed in all wells, then 5 µl, 10 µl or 20 µl of the virus was added to each well, leaving one well as a non-infected control. The next morning media was changed and 24 hours after virus addition, 10 µl of 0.1 mg/ml puromycin was added to every well to give a final concentration of 0.5 µg/ml. On day 2 the optimal volume of virus to provide efficient LUHMES cell infection could be approximated based on cell survival of puromycin treatment. All wells were trypsinised and centrifuged as described in section 2.1 and each entire well from the 6-well plate was seeded into a 10cm dish in proliferation media containing 0.25 µg/ml puromycin. On day 4 all of the non-infected WT LUHMES cells were dead, and samples that were efficiently infected with an appropriate amount of lentivirus will survive the puromycin selection. If samples were to be subjected to single-cell cloning, two days after they were moved to 10cm dishes, samples were taken to FACS sorting as described in section 2.4.

2.11 Genomic DNA extraction methods

For extracting the genomic DNA from tissue culture samples, Puregene Core Kit A (Qiagen) was used following manufacturer's instructions. Occasionally when RNA and DNA was required from a single sample, the Allprep DNA/RNA Mini kit (Qiagen) was used.

For large scale genomic DNA extraction from LUHMES single cell clones in a 24-well plate an alternative approach was used. Individual clones when confluent were trypsinised and centrifuged and incubated in 400 µl lysis buffer (50 mM Tris pH 9.0, 20 mM EDTA pH 8.0, 40 mM NaCl, 1% SDS, 0.5 mg/ml proteinase K) overnight at 55^{o}C. The next day 300 µl saturated NaCl was added to each sample, mixed by vigorous shaking for 1 minute and centrifuged at 14000 rpm for 10 minutes at room temperature. The supernatant was transferred to an Eppendorf containing 500 µl isopropanol, mixed by inversion and centrifuged at 14000 rpm for 10 minutes at 4^{o}C. The DNA pellet was washed with 750 µl 70% ethanol, mixed by inversion, and centrifuged at 14000 rpm for 10 minutes at 4^{o}C. The DNA pellet was air dried, resuspended in 50 µl TE and allowed to dissolve at 55^{o}C overnight.

2.12 RNA extraction, cDNA synthesis and qPCR analysis

RNA was extracted using RNeasy Mini Kit (Qiagen) following manufacturer's instructions. RNA concentration was measured using a Nanodrop 1000 (Thermo Scientific) and 10 µg RNA was taken for DNase1 treatment (Ambion) for 1 hour at 37°C. DNA-free RNA concentration was measured using a Nanodrop 1000 and 1 µg taken for cDNA synthesis using qScript cDNA Supermix (Quanta). DNA-free RNA was tested for the presence of genomic DNA by performing a PCR using GoTaq Polymerase (Promega) and primers for an intron-exon region within *GAPDH*. If cDNA was to be used for quantitative PCR analysis, the 20 µl cDNA reaction was diluted 10-fold using ddH_2O. The SensiMix SYBR and Fluorescein mix (Biolines) was used for qPCR reactions using a LightCycler480 machine (Roche). If qPCR was performed in a 96-well plate, 25 µl reactions were used, however if a 384-well plate was used, 12.5 µl reactions were used by halving the amount of each component as described in the Biolines protocol. Each sample for qPCR analysis was performed in technical triplicate and standard curves were made for every set of primers using 5 different dilutions. Analysis was performed as follows: using the standard curve the normalised Cp value of each sample was calculated, denoted X^{target}. The value of $10^{Xtarget}$ was calculated and the ratio of this for each experimental sample was compared to that of the control *GAPDH*: $10^{Xtarget}/10^{XGAPDH}$. This number was averaged for the technical triplicates. As samples were performed in biological triplicate (three neuronal differentiations that were set up at different times), the values shown in all graphs are the mean of these biological triplicates with the SEM plotted as error bars.

2.13 T7 Endonuclease I assay

Phusion polymerase in GC buffer (NEB) was used to PCR amplify 100-250 ng genomic DNA in a 50 µl reaction volume and 5 µl of the reaction was ran out on an agarose/TAE gel to confirm efficient amplification. To the remaining 45µl PCR mix, 5 µl of Buffer 2 (NEB) was added and the reactions were heated at 95°C for 10 minutes in a PCR machine and cooled slowly to 25°C to produce heteroduplexes. Each reaction was split in half and to one half 1µl of T7E1 (NEB) was added. All reactions were incubated at 37°C for 30 minutes, followed by addition of 10 µl of 6X loading dye (NEB) and analysis on an agarose/TAE gel.

2.14 Restriction fragment length polymorphism assay

Phusion polymerase in GC buffer was used to PCR amplify 100-250 ng genomic DNA in a 25 µl reaction volume and 10 µl of each PCR was run out on an agarose/TAE gel. To the

remaining PCR mix 29 µl of H_2O, 5 µl of CutSmart buffer (NEB) and 0.5 µl of the appropriate enzyme was added. Reactions were incubated at the necessary temperature and length of time as appropriate for each enzyme and were analysed by electrophoresis in an agarose/TAE gel.

2.15 PCR and sequencing analysis

Phusion polymerase in GC buffer was used to PCR amplify genomic DNA in a 25 µl reaction volume and 10 µl of the reaction was ran out on an agarose/TAE gel. Each amplicon was subcloned using the Strataclone blunt PCR cloning kit (Agilent Technologies) following the manufacturer's instructions and transformed into Strataclone Solopack competent bacteria (Agilent Technologies). The next day single colonies were picked and colony PCR was performed using either Phusion polymerase or DreamTaq polymerase in a 25 µl volume. For confirmation of PCR, 10 µl of each colony PCR reaction was analysed by agarose/TAE gel electrophoresis. Positive colony PCRs were treated with 0.25 µl Exonuclease I (NEB) and 0.25 µl FastAP alkaline phosphatase (Thermo) in a 28 µl total volume at 37^{o}C for 15 minutes followed by inactivation at 85^{o}C for 15 minutes. Sequencing was performed using BigDye Terminator v2.1 reaction mix (Life Technologies) using 3.5 µl of each sample and 3.2 pmol of primer. PCR reactions were sequenced by Edinburgh Genomics using Sanger sequencing.

2.16 Whole cell and nuclear extracts

Whole cell protein extracts were prepared by homogenising cell pellets in NE1 buffer (10 mM HEPES pH 7.9, 10 mM KCl, 1 mM $MgCl_2$, 0.5 mM DTT, 0.1% Triton-X 100, 20% glycerol and 1X protease inhibitor cocktail (Roche)). Homogenates were treated with benzonase for 15 minutes at room temperature and then measured for protein concentration using a Bradford assay (Protein Assay Dye Reagent concentrate, BioRad).

Nuclear extracts were prepared by resuspension of cell pellets into Buffer A (10 mM Tris pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% Igepal CA-630, 0.5 mM DTT, 1X protease inhibitor cocktail) and incubating on ice for 10-15 minutes. Cells were then dounced in a glass homogeniser (or for volumes smaller than 500 µl in a 1.5 ml Eppendorf with a plastic dounce) to release nuclei and nuclei were pelleted by centrifugation at 2400 rpm for 5 minutes at 4^{o}C. Nuclei were resuspended in NE1, and counted on a haemocytometer using

trypan blue staining if accurate nuclei concentrations were required. Nuclei were then treated with benzonase for 15 minutes at room temperature.

All samples were mixed with 2X Sample Buffer, Laemmli loading dye (S3401-1VL) prior to storage at -20°C and loading on SDS-PAGE gels.

2.17 SDS-PAGE, Coomassie staining and Western blot analysis

Samples were loaded onto pre-cast 4-20% Mini-PROTEAN TGX gels (BioRad) and ran at 200V for approximately 30-40 minutes. Gels were transferred onto nitrocellulose membrane by transfer at 25V overnight (~18 hours) at 4°C. A second membrane was used in the transfer in order to monitor excessive protein transfer. The gel was stained with a coomassie stain (InstantBlue, Expedeon) for 1 hour after transfer to determine the efficiency of transfer of proteins out of the gel. The two membranes were stained with Ponceau S for 2 minutes to monitor the transfer of proteins onto the 1st and 2nd membranes. Membranes were then de-stained with 0.1% Tween-20 in PBS, blocked in 5% milk, 0.1% Tween-20 in TBS for 30 minutes and probed with primary antibodies in blocking solution for 1 hour at room temperature (Section 2.2). Membranes were washed 3 times with 0.1% Tween-20/PBS. IRDye 800CW α-mouse and IRDye 680LT α-rabbit secondary antibodies (Licor) were probed for 1 hour at room temperature in blocking buffer at a concentration of 1:10000 and scans were taken using a Licor Odyssey machine.

2.17 Immunofluorescence imaging and microscopy

Cells grown on coverslips were fixed in 4% formaldehyde for 10 minutes, permeabilised with 0.2% Triton-X/PBS for 10 minutes and blocked for 30 minutes in 10% fetal bovine serum in PBS (FBS/PBS). Coverslips were incubated in the presence of primary antibodies (Table 2.5) in 1% FBS/0.1% Tween-20/PBS for one hour room temperature (section 2.2). Coverslips were washed with 0.1% Tween-20/PBS, and incubated for one hour at room temperature with Alexa Fluor secondary antibodies (1000X dilution, Invitrogen) in 1% FBS/0.1% Tween-20/PBS. Secondary antibodies: α-mouse 488, α-rabbit 555 and α-chicken 633. Coverslips were finally stained with DAPI (5000X dilution in PBS, Sigma, D9542) for 10 minutes at RT and mounted onto microscope slides using Prolong Diamond solution (Thermo Fisher). Z-stack images were taken using a Leica SP5 microscope and z-stacks were flattened and processed using ImageJ 1.47v software.

Phase contrast images were taken on an Eclipse TS100 microscope (Nikon) using QCapture Pro software version 5.1.1.14 (QImaging). UV fluorescence was used to take GFP pictures of cells. The IncuCyte microscope (Essen Biosciences) was used to take phase contract and GFP fluorescent images and ZOOM software (Essen biosciences) was used to extract and export the images.

2.18 Karyotyping

To karyotype LUHMES progenitor cells, LUHMES cells were cultured as described in section 2.2. The day before preparation of metaphase spreads, 0.5 x 10^6 LUHMES cells were seeded in a 6cm dish. The next day media was removed and fresh proliferation medium containing 0.1 µg/ml colcemid was added. After incubation for 3 hours, cells were washed twice with pre-warmed PBS, harvested by trypsinisation and resuspended in 300 µl medium with 5 ml 0.4% KCl. Cells were incubated in a 37ºC waterbath for 10 minutes then 100 µl of a 3:1 methanol:acetic acid mix was added to cells and mixed by gentle inversion. Cells were pelleted by centrifugation at 300 x g for 5 minutes and supernatant was removed. Cells were resuspended in 5 ml 3:1 methanol:acetic acid and incubated at room temperature for 20 minutes. Again cells were centrifuged at 300 x g for 5 minutes and supernatant was removed. The pellet was resuspended in 200 µl 3:1 methanol:acetic acid. This mixture was dropped onto pre-chilled microscope slides from an approximate 30 cm height and dried overnight at room temperature. The next day coverslips were mounted onto the slides using Vectorshield containing DAPI (Vectorlabs) and images were captured using a Zeiss AxioImager.Z1 microscope and counted using ImageJ 1.47v software.

2.19 DNA and RNA FISH

RNA and DNA FISH of LUHMES cells was performed in collaboration with Dr Ronan Chaligne and Prof Edith Heard at the Institut Curie in Paris. The Heard lab provided me with paraformaldehyde (PFA) and VRC (RNase inhibitor) which were stored at -20ºC prior to use. LUHMES cells were cultured on Nunc Lab-Tek II chamber slides (Thermo, 154453) that contained 1 well and were provided with a lid for a sterile tissue culture environment. Proliferating LUHMES cells were seeded at a density of 0.36 x 10^6 cells/slide and fixed 2 days later, or at a density of 0.6 x 10^6 cells/slide and fixed the following day. LUHMES cells on day 2 of differentiation were seeded at a density of 1.2 x 10^6 cells/slide and fixed on day 6 of differentiation. On the day of fixation, cell culture medium was aspirated and cells were washed once with PBS. At this stage the plastic cover surrounding the glass slide was removed to allow fixation and permeabilisation on the naked slides in plastic Coplin jars

(Heathrow Scientific, HS15986). Cells were fixed for 10 minutes at room temperature in 3% PFA/PBS, followed by permeabilisation in pre-chilled 0.5% Triton-X/1% VRC/PBS for 5 minutes on ice. Cells were then washed once in PBS, twice in 70% ethanol/DEPC-treated water, and stored in the plastic Coplin jars in 70% ethanol at -20ºC. Slides were then shipped to Paris in these jars and surrounded by ice blocks that had been cooled to -20ºC. Dr Chaligne then performed DNA and RNA FISH as described in (Chaumeil *et al.*, 2008). BAC-based probes were used for detection of *MECP2* and *ATRX* (RP11-119A22 and RP11-42M11 respectively). Probes for *XIST* RNA detection have been used previously, as described in (Chaligne *et al.*, 2015).

2.20 Nuclear Volume measurements

Neurons were grown on 19 mm coverslips until day 9 and then fixed and stained against laminB1 as described in section 2.17. Z-stack images were taken on a Leica SP5 confocal microscope and analysed using ImagePro Premier software. The software identified spheres of laminB1 staining and calculated the volume within this. I manually went through each image of nuclei spheres and altered the size range for each image to exclude laminB1 staining of small debris and adjacent nuclei that were counted as one.

2.21 Calcium imaging

For all calcium imaging experiments, neurons were grown on 25 mm coverslips, thickness #1.5 (Warner Instruments, 64-0715) in 6-well plates. At the desired age, neurons were washed twice with HBSS lacking calcium and magnesium (Lonza, 10547F) supplemented with 10 µM HEPES and then loaded with the required loading solution. For trials of calcium imaging, the loading solutions are described in Table 7.

The incubation time for Fluo-4 samples was 1 hour at 37ºC, while the incubation of Rhod-3 samples was at room temperature for 1 hour. After Rhod-3 loading buffer incubation, coverslips were washed twice with HBSS+HEPES and then incubated for a further hour at room temperature in HBSS+HEPES+2.5 mM probenecid. After incubation, coverslips were washed once with HBSS+HEPES and then an HBSS containing calcium and magnesium (Lonza, BE10527F) + 10 µM HEPES solution was added to the cells, plates were wrapped in foil and taken for imaging.

1mM Fluo-4 stock (µl)	100X Powerload (µl)	250mM Probenicid (µl)	HEPES (ml)	Final Fluo-4 concentration (µM)
4	-	-	1	2
7	-	-	1	3.5
10	-	-	1	5
4	20	20	0.976	2
7	20	20	0.973	3.5
10	20	20	0.970	5
10mM Rhod-3 stock (µl)	**100X PowerLoad (µl)**	**250mM Probenecid (µl)**	**HEPES (ml)**	**Final Rhod-3 concentration (µM)**
4	20	20	0.976	20
4	-	-	1	20
2	20	20	0.978	10
2	-	-	1	10
1	20	20	0.979	5
1	-	-	1	5

Table 7. Volumes and concentrations of components in the loading buffer when trialling Rhod-3 and Fluo-4 calcium fluorophores.

100X PowerLoad and probenecid were provided in the Rhod-3 calcium imaging kit (Life Technologies, R0145). Fluo-4 (Invitrogen, F14201) and Rhod-3 were reconstituted using DMSO provided in the Rhod-3 kit.

Plates containing coverslips in HBSS were kept in a 37°C incubator. One by one coverslips were placed into a coverslip holder and 200 µl HBSS+Ca^{2+}+Mg^{2+}+HEPES was added to the cells and placed on the Leica SP5 microscope with a 40X objective. Once neurons were in focus, filming began and after 20 frames of imaging "resting neurons", 200 µl of solution A was added to the coverslip and neurons were imaged for a further 500 image frames. Note that 1 frame is the equivalent of 0.65 seconds; therefore the total imaging time for a single field of view was 338 seconds, or 5 minutes and 38 seconds. For a negative control, solution A was 2X EGTA (Sigma-Aldrich, E8145); for a positive control, solution A was 2X A23187 calcium ionophore (Fisher Scientific, 10284963); and for experimental studies, solution A was either 130mM or 260mM KCl (Fisher Scientific - P/4280/6).

After trials, all experiments were performed using 2 µM Fluo-4 were loaded for 1 hour at 37°C with Fluo-4 in DMSO, Pluronic-F127 (Biotium, 59004), HBSS and 10 µM HEPES. For example, for 6 coverslips a total of 12ml of loading buffer is needed. This will consist of 12 ml HBSS+HEPES and 32 µl of 1 mM Fluo-4, where the 32 µl consists of 16 µl of Fluo-4 in DMSO and 16 µl of Pluronic-F127. Note each 50 µg vial of Fluo-4 was freshly

reconstituted in 23 µl DMSO and any reconstituted Fluo-4 that was not used was returned to -20^{o}C for storage and would only be thawed once.

Analysis was performed by picking a minimum of 6 different cell bodies per video, and defining the recording area as a circle encompassing the total cell body area. The mean fluorescence intensity in this area was calculated for the duration of the time course and this was averaged among the 6 replicates. The maximum fluorescence intensity from this averaged time course was defined and this was plotted with respect to the calcium concentration. The K_d for Fluo-4 was then established by reading the calcium ion concentration for the 50% fluorescence intensity from the graph.

2.22 HPLC

High performance liquid chromatography was performed in collaboration with Dr Bernard Ramsahoye in order to measure total RNA and DNA amounts. LUHMES cells were proliferated in a 6-well plate until confluent or differentiated in 6-well plates and harvested on day 9. One well of a 6-well plate provided enough material for undifferentiated samples, but neuronal samples required two wells of a 6-well plate to be pooled. Cells were washed once with PBS, then scraped in 500 µl ice-cold lysis buffer (10 mM Tris-HCl pH 7.5, 100 mM EDTA) and transferred into round-bottomed 2 ml Eppendorf tubes on ice. Once all samples had been harvested, 7.5 µl of 20 mg/ml Proteinase K and 12.5 µl of 20% SDS was added to each sample and mixed by flicking and inversion. Samples were put to 55^{o}C for 1 hour before being transported to Dr Ramsahoye at room temperature for HPLC analysis. 11-week old R306C male mice and wild-type littermate's brain sections were harvested by Dr Jim Selfridge and stored in liquid nitrogen as a whole hypothalamus, ½ cortex and ½ cerebellum samples. Each mouse brain section was homogenised using a glass hand-held homogeniser in 400 µl ice-cold lysis buffer containing 300 µg/ml Proteinase K using approximately 10 strokes and then stored on ice. Once all samples had been homogenised, 10 µl 20% SDS was added to each sample and mixed by inversion. Samples were incubated at 55^{o}C for 30 minutes and then transported to Dr Ramsahoye at room temperature for HPLC analysis.

2.23 RNA-sequencing

LUHMES cells were differentiated in 6-well plates to day 9 with three wells being pooled to form one sample for RNA-seq. The Allprep DNA/RNA Mini Kit (Qiagen) was used to

extract total RNA and DNA from each sample eluting in 40 µl and 100 µl respectively. RNA and DNA concentrations were measured using a Qubit fluorometer (Invitrogen) using the RNA broad-range and the DNA high-sensitivity kits for RNA samples and the DNA broad-range and RNA high-sensitivity kits for RNA samples. After removal of 1 µl of sample, to the 39 µl remaining RNA sample, 5 µl buffer, 5 µl H_2O and 1 µl rDNaseI (Ambion) was added and samples were incubated at 37^{o}C for 40 minutes to digest the DNA. RNA samples were again measured on the Qubit using the RNA broad-range and DNA high-sensitivity kits. A PCR to check for zero genomic DNA contamination was performed using 1 µl of each sample (plus the 1 µl of each sample prior to rDNaseI treatment) using GoTaq polymerase (Promega) and primers crossing an intron-exon boundary of *GAPDH*. RNA samples were then assessed for quality using a RNA Nano 6000 chip on the Bioanalyser 2100 machine (Agilent).

The ScriptSeq Complete kit for human/mouse/rat (Epicentre) was used for removal of rRNA, fragmentation of RNA, preparation of cDNA and PCR amplification by following the manufacturer's instructions. Briefly, magnetic beads for rRNA removal were washed in batches of 6 using gentle vortexing and were stored at room temperature with 1 µl RiboGuard RNase Inhibitor until needed. To 2.5 µg of DNA-free RNA, Ribo-Zero solutions were added in a reaction that included 5 µl of 100X ERCC spike-in mix (Thermo Fisher); components were added individually to each reaction. Samples were then applied to the prepared magnetic beads and the supernatant from this step was cleaned up using an Agencourt RNAClean XP kit (Beckman Coulter). The mRNA samples were then tested for quality using an mRNA Pico chip on the Bioanalyser and were quantified using the Qubit fluorometer and a RNA high-sensitivity kit. At this stage mRNA samples were frozen at -80^{o}C. For fragmentation of RNA, 8 ng of mRNA was used. Synthesis of cDNA was then performed, 3' terminal tagging and purification of cDNA using AMPure XP beads (Beckman Coulter). Library amplification was then performed using the forward primer in this ScriptSeq Complete kit and reverse primers from the ScriptSeq Index PCR primers kit (EpiCentre) in order to index the individual samples and 13 cycles of PCR amplification. PCR samples were then cleaned up using AMPure XP beads and library quality was assessed using a high sensitivity DNA chip on the Bioanalyser machine before being sent off for sequencing at the Sanger institute in Cambridge using Illumina HiSeq 2500 paired-end sequencing on 4 lanes (equivalent to 2 rapid runs, 6 samples per run).

2.24 RNA-sequencing analysis

RNA-sequencing analysis was performed by Kashyap Chaatbar as follows. Poor quality raw sequencing reads are removed and the remaining reads are adapter trimmed using Trimmomatic v0.33 (Bolger *et al.*, 2014). The remaining reads are aligned to the hg19 version of Human genome assembly (Consortium, 2001) using STAR v2.4.2a (Dobin *et al.*, 2013). Genomic features are taken from Ensembl release 74 (Yates *et al.*, 2016). The number of reads mapping to each genomic feature is quantified using featureCounts v1.5.0 (Liao *et al.*, 2014). After the quantification of the mapped reads, differential gene expression is calculated using DESeq2 v1.10.1 (Love *et al.*, 2014). The command line arguments used to perform these analyses is as follows:

```
java -jar trimmomatic-0.33.jar PE ILLUMINACLIP:TruSeq2-PE.fa:2:30:10 LEADING:3
TRAILING:3 SLIDINGWINDOW:4:30 MINLEN:30

STAR --runThreadN 10 --runMode alignReads --genomeDir hg19 --outFilterIntronMotifs
RemoveNoncanonicalUnannotated --outSAMtype BAM SortedByCoordinate

featureCounts -T 10 -p -s 1 --primary -t exon -g gene_id -a hg19.gtf
```

Calculation of mCpG densities were performed as follows:
mCpG density (ρ) and mCpG mean ($\underline{x}$) calculation representation

$$\rho = \frac{\Sigma \frac{\textbf{Number of methylated reads on a particular CpG}}{\textbf{Total number of reads at that CpG}}}{\textbf{Length of region (kb)}}$$

$$\underline{x} = \frac{\Sigma \frac{\textbf{Number of methylated reads on a particular CpG}}{\textbf{Total number of reads at that CpG}}}{\textbf{Number of CpG in the region}}$$

Gene length, promoter length and gene body length were calculated as follows: Gene features for human genome assembly GRCh37 were taken from Ensembl version 74 and represent regions between the annotated TSS and TES. Promoters were defined as 2kb regions centred around the TSS. To create gene body annotations we first removed all genes with length < 3.5kb. Gene bodies define regions that start 3kb downstream of the TSS and

lead to the TES. In other words: Gene = Annotated TSS to TES; Promoter = 1kb either side of TSS; Gene body = Remove genes < 3.5kb in length, take 3kb downstream of TSS to TES.

2.24 IncuCyte measurements

Long-term image analysis of differentiating and maturing LUHMES neurons was performed in an IncuCyte machine (Essen Biosciences). Neurons were differentiated in both 6-well and 96-well plates (TPP, Sigma) and every well was imaged at 4 time points during a 12-hour period. Images were analysed using the ZOOM software (Essen Biosciences) with the NeuroTrack software able to measure the length of neurites during the differentiation time course. The All Older processing definition was designed in this project. The parameters for the All Older processing definition are as follows; segmentation mode brightness, segmentation adjustment 1.1, neurite filtering best, neurite sensitivity 0.35 and neurite width 1 µm.

2.25 Electrophysiological measurements

LUHMES neurons were grown on 14 mm coverslips (Thermo) and transported in a polystyrene box heated to 37^{o}C to Dr Sean McKay who performed patch-clamping experiments. Coverslips were transferred to a recording chamber perfused (at a flow rate of 3–5 ml/min) with an external recording solution composed of (in mM): 150 NaCl, 2.8 KCl, 10 HEPES, 2 CaCl2, 1 MgCl2, and 10 glucose, pH 7.3 (320–330 mOsm). Patch-pipettes were made from thick-walled borosilicate glass (Harvard Apparatus), and when filled with the internal recording solution had tip resistances of 4–8 MΩ. A K-gluconate-based internal solution was used for patching neurons composed of (in mM): 141 K-gluconate, 2.5 NaCl, 10 HEPES, and 11 EGTA; pH 7.3 with KOH. Recordings were initially made in voltage-clamp and rejected if the holding current was greater than −100 pA or if the series resistance drifted by more than 20% of its initial value (< 30 MΩ) by the end of the recording. During the current-clamp recordings, no current was injected at the resting membrane potential and neuronal firing was induced by current injections from -20pA to +50pA in incremental 5pA steps. Input resistance was significantly lower than primary mouse cultures recorded the same day (data not shown). Recordings were at room temperature (21 ± 2 °C) using a Multiclamp 200B amplifier (Molecular Devices). Recordings were filtered at 5 kHz and digitized online at 20 kHz via a BNC- 2090A/PCI-6251 DAQ board interface (National Instruments) and analysed using WinEDR 3.6 software (Dr John Dempster, University of Strathclyde).

2.26 Protein transduction

Wild-type human MeCP2_e1 isoform MeCP2 was made by Dr Matthew Lyst as described previously (Klose & Bird, 2004) with some minor alterations; BL21 codon plus bacteria were swopped for pLysS, Tris-HCl was exchanged for HEPES in the buffers, NP-40 was exchanged for TritonX-100, SP-sepharose columns were performed prior to Ni-NTA, and no Sephacryl S-300 26/60 or MonoS column purifications strategies were performed. MeCP2_e1 protein was stored in aliquots at -20^{o}C. Chloroquine (Sigma, C6628-25G) was prepared fresh by dissolving 0.1 g of powder in 1 ml tissue-culture grade PBS (Gibco, 14190).

Wild-type and MeCP2 KO (2_7 cell line) LUHMES-derived neurons were differentiated in 6 cm dishes or on coverslips. Each experiment contained 5 samples: WT cells, KO cells, KO cells + choloroquine, KO cells + protein, KO cells + chloroquine + protein. Human On day 9 of differentiation, media was removed from all dishes, mixed in a 1:1 ratio with fresh LUHMES differentiation medium and 4 ml aliquots were made. To each aliquot the required amount of protein and/or chloroquine was added (Table 2.8), mixed by inversion and added to neurons. A media change was also performed for WT or KO neurons that were not treated with protein or chloroquine. After the desired incubation time neurons were either fixed for immunofluorescence analysis, or harvested by washing once in tissue-culture grade PBS, scraping in ice-cold PBS and centrifugation at 300 x g for 5 minutes at 4^{o}C. Cell pellets were snap-frozen on dry ice and stored at -80^{o}C.

	Old protein prep 6cm dish	Old protein prep Coverslip	New protein prep 6cm dish	New protein prep 24-well plate
Cells seeded	2 x 10^6	0.15 x 10^6	2 x 10^6	0.15 x 10^6
Media volume	2 ml	1 ml	2 ml	0.5 ml
Chloroquine (100 mM)	2 µl	0.5 µl	4 µl	0.1 µl
MeCP2 (3.8 or 1 mg/ml)	25 µg (6.58 µl)	1.9 µg (0.5 µl)	2.5 µg (2.5 µl)	0.125 µg (0.125 µl)
MeCP2/cell	12.5 pg/cell	12.6 pg/cell	1.25 pg/cell	0.83 pg/cell

Table 8. Amounts and final concentrations of MeCP2_1e protein and chloroquine for protein transduction.

Amounts that were added to LUHMES-derived neurons in 6 cm dishes, on coverslips in 12-well plates and in 24-well plates. Note, 6-day old LUHMES neurons have approximately 0.256 pg MeCP2/nucleus.

Chapter 3

3. Characterisation of the LUHMES neuronal cell line and its utility as a RTT model system

3.1 Introduction

The LUHMES human, pre-neuronal cell line was established in 2005 (Lotharius *et al.*, 2005). There has been extensive RT-qPCR and immunofluorescence analysis of the differentiation dynamics of the cells which demonstrate the neuronal-committed nature of the progenitors and their relatively homogeneous differentiation towards a dopaminergic lineage (Scholz *et al.*, 2011). Furthermore, a quick and efficient differentiation protocol has been described (Scholz *et al.*, 2011), and overexpression of proteins in progenitors via lentiviruses and siRNA knock-down of protein in 2-day old pre-differentiated cells via Lipofectamine-2000 transfection has been demonstrated, indicating an ability to manipulate these cells (Schildknecht *et al.*, 2013). The ease of differentiation, the ability to manipulate and the homogeneous nature of this human neuronal cell line make them an attractive tool for the study of MeCP2 and Rett syndrome.

Much of our MeCP2-focussed work has involved whole organism studies of mutant mouse lines combined with transfection experiments in mouse NIH3T3 or human HeLa cell lines. The LUHMES cell line can potentially provide a half-way point between these two. The overall aim of this project is to genetically manipulate the endogenous *MECP2* gene in LUHMES cells in order to create human neurons where the endogenous protein can be studied in its native context. I aim to show that LUHMES cells can be handled in a high-throughput manner and that there are a range of experiments that can be performed leading to novel discoveries. I do not wish to propose that LUHMES cells will be an alternative to mouse models and work-horse cell lines, but I do believe that they will complement these

studies by providing an easy-to-manipulate, human neuronal environment for study of the MeCP2 protein.

In order to establish the LUHMES cell line as a model system for the laboratory, it is first necessary to perform some basic characterisations of these cells. Are LUHMES cells an easy to handle and karyotypically normal cell line? Do they express high levels of MeCP2, as would be expected of a neuronal cell line? Can they be transfected at the progenitor stage in order to produce stocks of genetically modified cell lines? Are they amenable to basic electrophysiology experiments that could be useful for downstream analysis if mutant cell lines can be generated? This chapter will address these questions.

3.2 Efficient and homogeneous differentiation of LUHMES cells into dopaminergic neurons

3.2.1 Differentiation timeline and phase contrast pictures

LUHMES cells were established as an immortalised cell line by infecting midbrain-isolated neuronal progenitor cells with a retroviral vector, called LINXv-myc, that expresses the *v-myc* oncogene from an altered CMV promoter (Hoshimaru *et al.*, 1996). The promoter (officially termed $P_{hCMV*-1}$) is a minimal promoter from human cytomegalovirus that is fused to tetracycline operon sequences (Gossen & Bujard, 1992). Upstream of this $P_{hCMV*-1}$-*v-myc* element is an expression construct for a tetracycline-controlled transactivator that acts upon $P_{hCMV*-1}$ to induce high levels of transcription in the absence of tetracycline. Upon addition of tetcracyline to the cell culture medium, the drug binds to the transactivator, inhibits its ability to regulate $P_{hCMV*-1}$ (Gossen & Bujard, 1992), and shuts off *v-myc* expression, thus allowing exogenous control over endogenous *v-myc* expression. Once *v-myc* is down-regulated, the LUHMES progenitors rapidly differentiate into a population of neurons, according to their cell fate and differentiation potential (Figure 3A).

Figure 3B illustrates the differentiation time course and rapid changes in morphology upon tetracycline addition to the cell culture media. Two days after tetracycline addition, the cells are expected to have shut down *v-myc* expression and therefore on this day cells are trypsinised and counted in order to seed a known number of cells into new vessels and thus control for cell number during experiments.

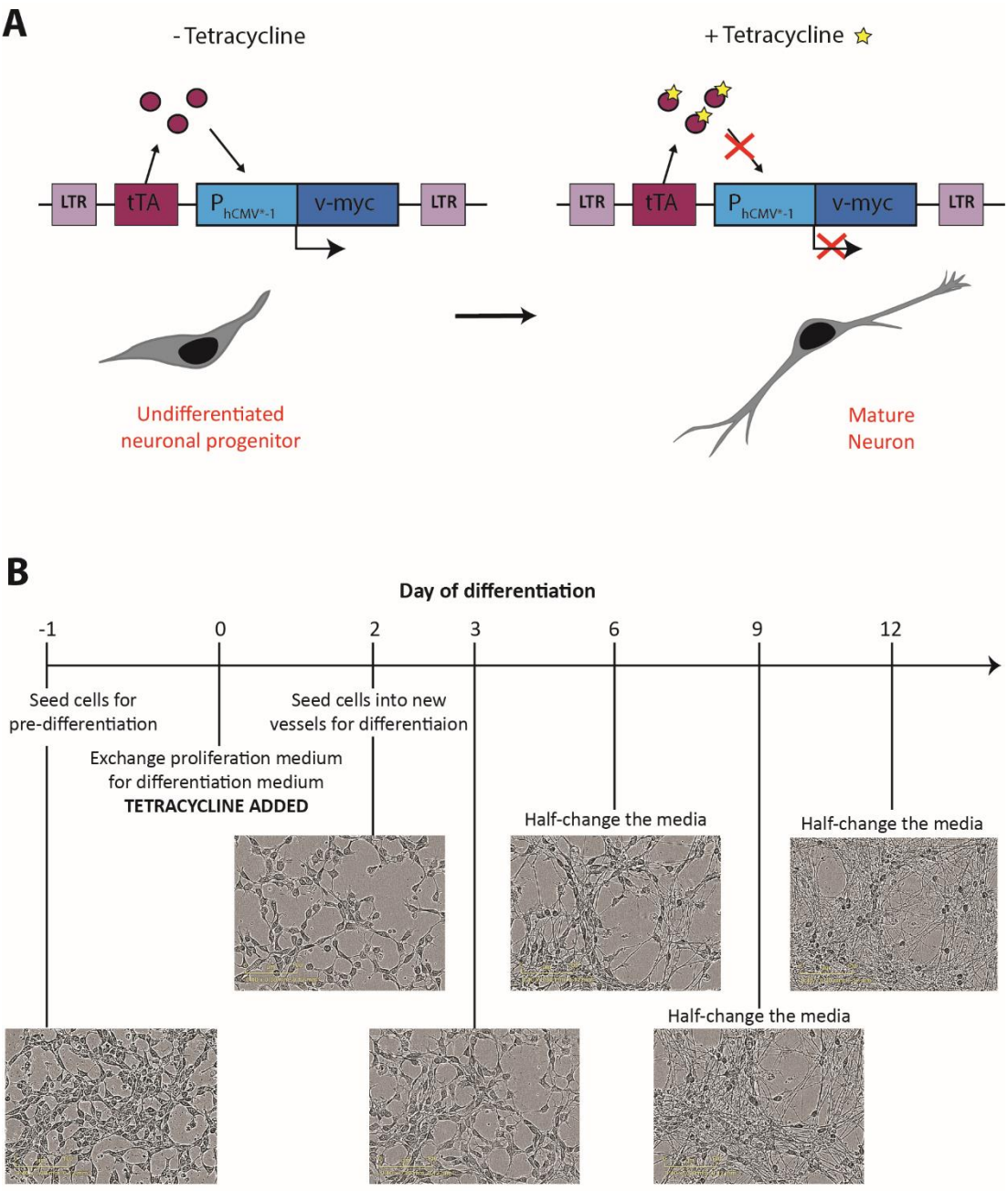


Figure 3. LUHMES cells show morphological changes upon addition of tetracycline indicative of a neuronal differentiation process.

(A) Schematic illustrating the principle behind LUHMES cell differentiation into neurons. The immature LUHMES neuronal progenitors have been immortalised with a retroviral vector such that they constitutively express *v-myc*. This *v-myc* expression is controlled by a TET-off system, thus simple addition of tetracycline to the cell culture media allows LUHMES cells to differentiate into neurons. **(B)** Timeline and representative phase contrast images of LUHMES cells as they differentiate from neuronal precursors into mature post-mitotic neurons. Scale bar is 100 µm.

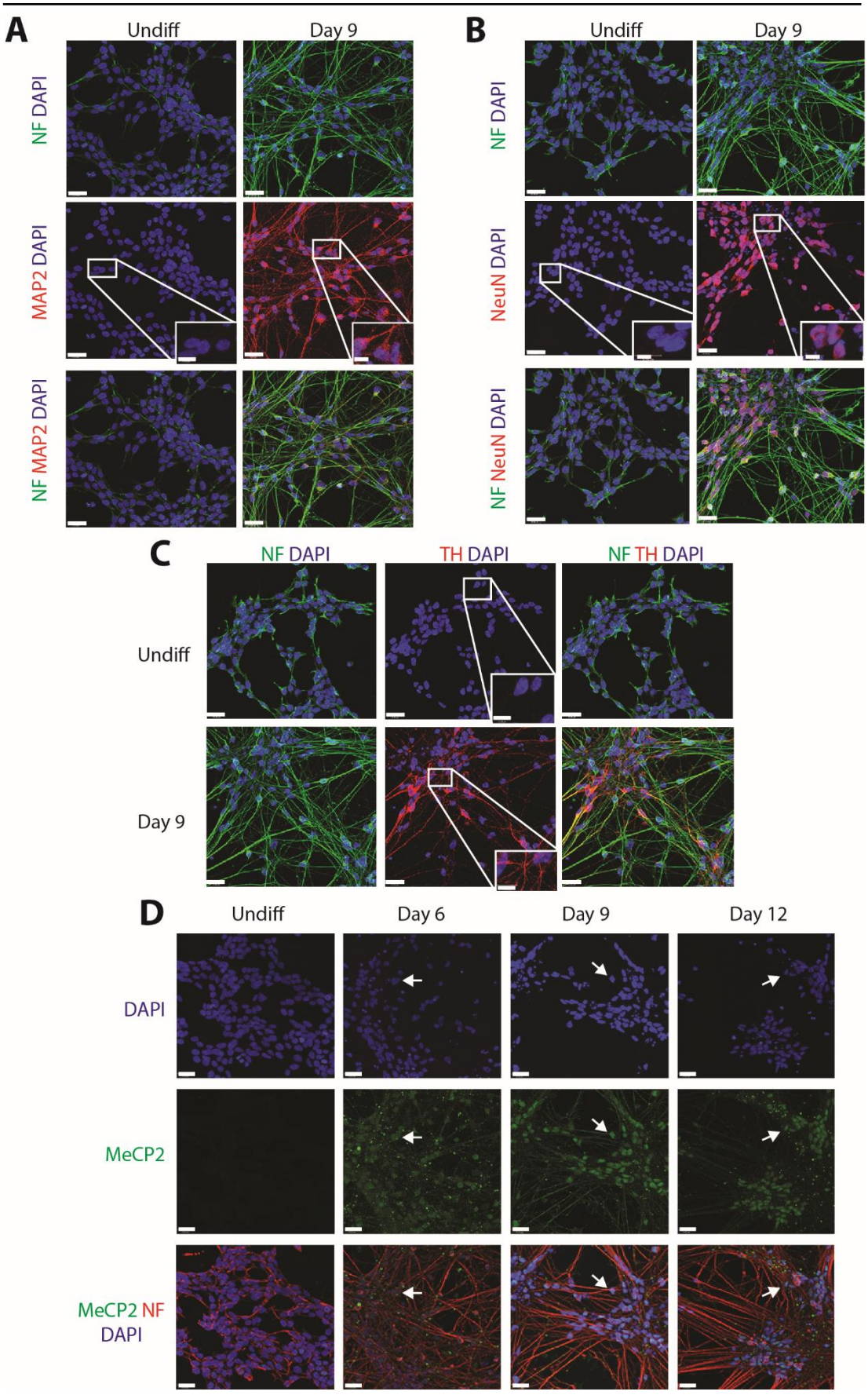

A
Undiff
Day 9
NF DAPI
MAP2 DAPI
NF MAP2 DAPI
B
Undiff
Day 9
NF DAPI
NeuN DAPI
NF NeuN DAPI
C
NF DAPI
TH DAPI
NF TH DAPI
Undiff
Day 9
D
Undiff
Day 6
Day 9
Day 12
DAPI
MeCP2
MeCP2 NF
DAPI

Figure 4. Immunofluorescence analysis demonstrates the differentiation of LUHMES cells into mature neurons within nine days.

(A) Staining of neurofilament (NF) and MAP2 in LUHMES cells, and in 9-day old LUHMES-derived neurons. **(B)** Staining of NF and NeuN in LUHMES cells, and in 9-day old LUHMES-derived neurons. **(C)** Staining of NF and tyrosine hydroxylase (TH) in LUHMES cells, and in 9-day old LUHMES-derived neurons. **(D)** Staining of NF and MeCP2 in undifferentiated LUHMES, and at days 6, 9 and 12 of differentiation. All images are stained with DAPI to identify the nucleus. Scale bars are 45 µm for main images, and 12 µm for inserts.

3.2.2 Immunofluorescence analysis of neuronal markers

Analysis of the expression of various markers of mature neurons shows that 9-day old LUHMES-derived neurons express high levels of MAP2, NeuN and neurofilament (Figure 4A+B). Indeed even the progenitor cells express neurofilament at low levels indicating their commitment to a neuronal lineage and absence of a stem cell-like, pluripotent state. Expression of tyrosine hydroxylase (TH) is observed at day 9, indicating the dopaminergic lineage of the neurons, (Figure 4C) and MeCP2 expression is gradually up-regulated during the differentiation process, characteristic of *in vivo* neuronal differentiation profiles (Figure 4D).

3.2.3 Quantitative PCR analysis of the dynamics of neuronal differentiation

In order to assess the dynamics of LUHMES cell differentiation into neurons, I performed RT-qPCR analysis of 25 different mRNAs, including markers of proliferating cells, immature neurons, mature neurons, and dopaminergic neurons, with the aim of establishing the dynamics of neuronal differentiation. Many primer pairs were provided by Dr Justyna Cholewa-Waclaw, whereas others were designed anew. To test the integrity of these new primer pairs melt curve analysis was performed (Supplementary Figure 1). Three primer pairs unfortunately consistently produced secondary products during the amplification phase of the qPCR reaction at three different annealing temperatures as indicated by melt curve analysis (Supplementary Figure 1A) and by running the final qPCR samples on an agarose gel (Supplementary Figure 1B+C). The remaining 10 primer pairs, however, produced clean melt curves indicating efficient amplification of a single product during the qPCR reaction (Supplementary Figure 1D).

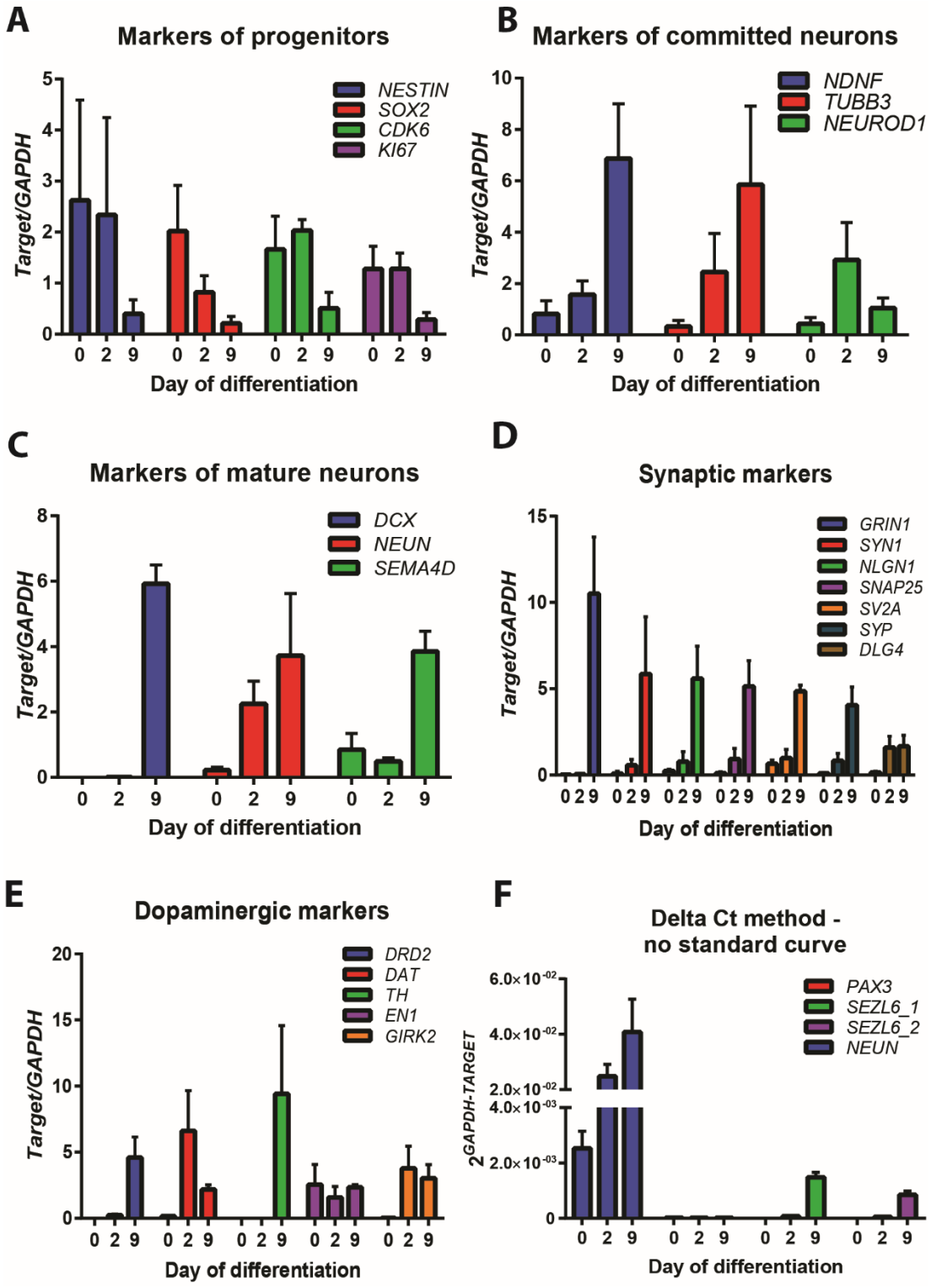
Figure 5 shows the qPCR analysis for all 25 mRNA species assessed. Markers of progenitor cells rapidly decrease as differentiation progresses (A), while markers for committed (B) and mature neurons (C) increase during differentiation. Seven markers of synaptic proteins (D) also increase their expression during differentiation and four markers for dopaminergic neurons (E) are up-regulated indicating the differentiation of LUHMES progenitors into mature, dopaminergic neurons. One dopaminergic marker, *EN1* (engrailed 1), does not upregulate its expression during differentiation, but remains expressed at all ages, consistent with its role as a transcription factor that functions throughout the development of midbrain dopaminergic neurons (Rekaik *et al.*, 2015). *PAX3* mRNA, a transcription factor found in migrating neuroblasts, could not be detected at any stage of differentiation (

Figure 5F). While previous analysis of *PAX3* could detect mRNA at days 0 and 5 of differentiation, no PAX3 protein was detected (Scholz *et al.*, 2011). The data thus indicates that this protein is most likely not expressed in LUHMES cells or their derived neurons. *SEZL6* is a transmembrane protein whose molecular function is unknown but is thought to involve cell surface signalling (Pigoni *et al.*, 2016). Using two different sets of primers, *SEZL6* could not be detected in proliferating or immature neurons, but a small level of expression (almost equivalent to NeuN expression in undifferentiated cells) was detected in 9-day old neurons (

Figure 5F). These qPCR experiments demonstrate the fast dynamics of LUHMES cell differentiation into mature neurons expressing a complement of transcription factors (*EN1*, *NEUROD1*), synaptic proteins (*SYN1*, *SNAP25*, *SV2A*), structural proteins (*TUBB3*, *NESTIN*), signalling proteins (*NDNF*, *DCX*) and cell surface proteins (*DRD2*, *DAT*, *GIRK2*, *GRIN1*) that are required for proper neuronal maturation and function.

Figure 5. Rapid differentiation into mature, dopaminergic neurons as determined by quantitative PCR analysis during LUHMES cell differentiation.

RT-qPCR analysis of wild-type undifferentiated LUHMES cells (day 0), 2-day old neurons and 9-day old neurons. **(A)** Proliferating, **(B)** committed, **(C)** mature, **(D)** synaptic transmission and **(E)** dopaminergic neuronal markers are shown relative to GAPDH expression. **(F)** No standard curve could be generated for *PAX3* and *SEZL6*. As such the delta Ct method of relative quantification to *GAPDH* was performed for analysis. *NEUN* is shown for a comparison to an mRNA species that increases its expression during LUHMES cell neuronal differentiation. *SEZL6* qPCR was performed using two different primer pairs (1 and 2).


A
Markers of progenitors
NESTIN
SOX2
CDK6
KI67
Target/GAPDH
Day of differentiation
B
Markers of committed neurons
NDNF
TUBB3
NEUROD1
Target/GAPDH
Day of differentiation
C
Markers of mature neurons
DCX
NEUN
SEMA4D
Target/GAPDH
Day of differentiation
D
Synaptic markers
GRIN1
SYN1
NLGN1
SNAP25
SV2A
SYP
DLG4
Target/GAPDH
Day of differentiation
E
Dopaminergic markers
DRD2
DAT
TH
EN1
GIRK2
Target/GAPDH
Day of differentiation
F
Delta Ct method -
no standard curve
PAX3
SEZL6_1
SEZL6_2
NEUN
2GAPDH-TARGET
Day of differentiation

3.3 Analysis of genomic integrity of LUHMES cells

3.3.1 Karyotyping

The LUHMES cell line is derived from healthy, wild-type tissue and might be predicted to have a stable karyotype. The LUHMES-precursor cell line, MESC2.10 cells, contain a normal complement of chromosomes (Paul *et al.*, 2007), but as this has not been demonstrated before in LUHMES cells themselves I chose to perform karyotyping. I generated metaphase spreads of LUHMES progenitors (Figure 6A) and counted the number of chromosomes per nucleus. The quantification shown in Figure 6B shows a broad distribution of chromosome numbers. This broad distribution can be accounted for by the presence of nuclei that had lost one or even three chromosomes from their spread, and therefore would be counted as 45 and 43 in the analysis (Figure 6A). Furthermore when two nuclei were overlapping or very close together it became difficult to accurately quantify the number of chromosomes in each spread (Figure 6A). This analysis was performed once, but with a median chromosome count of 46, I inferred a normal karyotype for LUHMES cells.

3.3.2 RNA and DNA FISH

Part of this study will deal with CRISPR targeting experiments of the X-linked *MECP2* gene and as LUHMES cells are female cells with two X chromosomes it was important to determine if both X chromosomes were active or if, as might be expected, one X chromosome was already established in the inactive state in progenitor LUHMES cells. I fixed and permeabilised progenitor cells as well as 6-day old neurons and Dr Ronan Chaligné (Prof Edith Heard's lab, Institut Curie, Paris, France) performed RNA and DNA FISH experiments using *ATRX* and *FMR1* as X-chromosome located controls. As shown in Figure 6C, a cloud denoting *XIST* RNA coating the inactive X chromosome was observed in progenitor cells and 6-day old neurons. A single focus of *MECP2* RNA expression was observed in the undifferentiated sample thus establishing that X inactivation is present in the progenitor cell line. *MECP2* is reported to escape X chromosome inactivation in trophoblast giant cells (Corbel *et al.*, 2013), in the subventricular zone *in vivo* (Gendrel *et al.*, 2014), and in mouse embryonic stem cells that are differentiated into neural precursors (Prof Edith Heard, personal communication). We observed 4% of LUHMES progenitors and 6-day old neurons expressing *MECP2* from the inactive X chromosome (Figure 6D and white arrowhead in Figure 6C). This percentage is much reduced compared to published reports of this phenomenon (30-40% in TGCs and 20% in the SVZ) and thus I conclude that the *MECP2* locus does not exhibit bi-allelic expression in LUHMES cells nor in LUHMES-derived neurons. Finally, DNA FISH confirmed the presence of two alleles of *MECP2*

genomic DNA in LUHMES progenitors, one of which co-localised with a cloud of *XIST* RNA expression (Figure 6E), thus providing further evidence for a stable genotype in this human cell line. To conclude, LUHMES cells have only two copies of the *MECP2* allele in their genome, and X chromosome inactivation is established in the progenitor cell line.

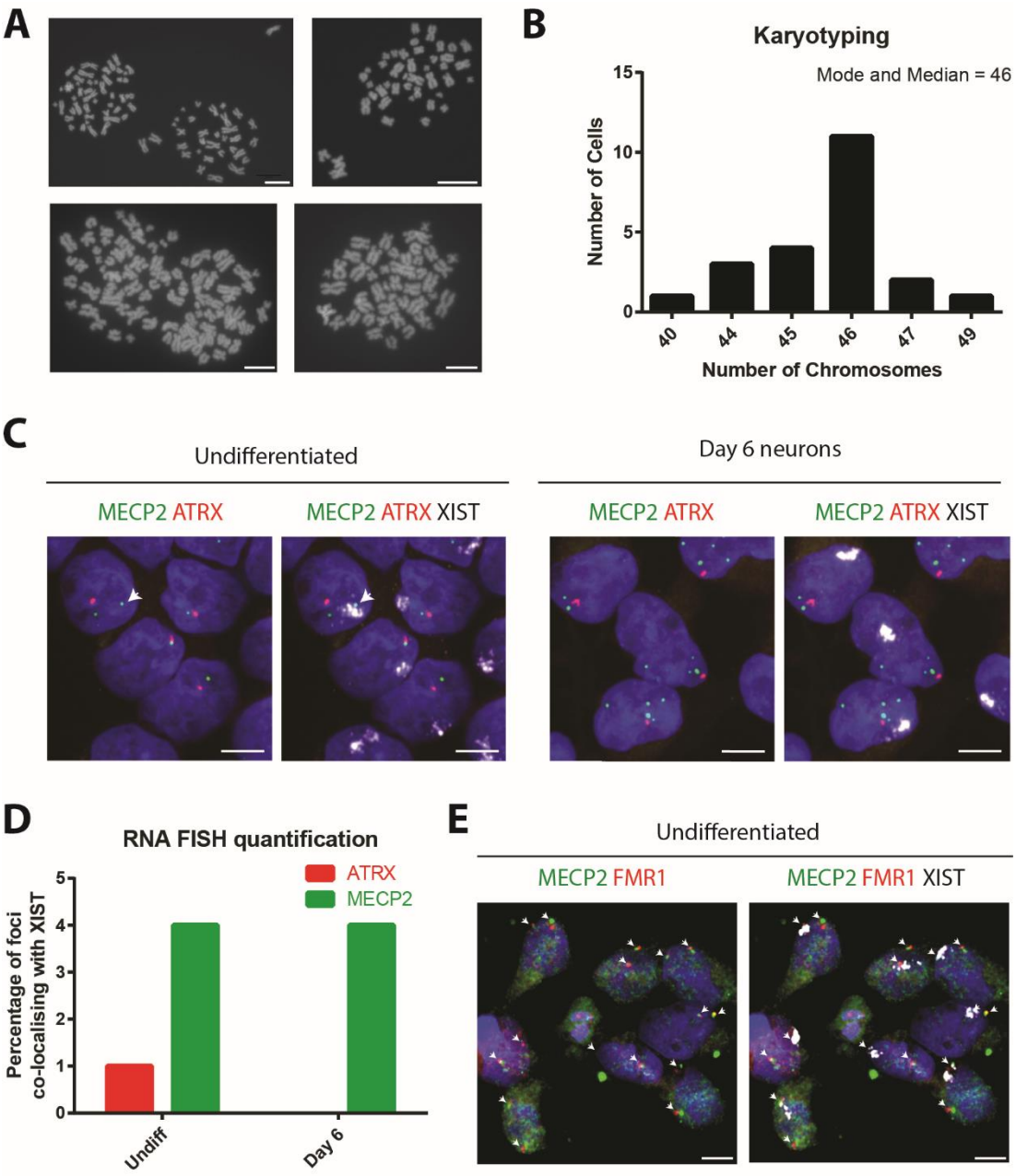


Figure 6. Assessing the genomic integrity of LUHMES cells and LUHMES-derived neurons.

(A) Representative images of chromosomes in a metaphase spread. **(B)** Quantification of the number of chromosomes per undifferentiated LUHMES cell. **(C)** RNA FISH analysis of LUHMES cells in the proliferative state and after 6 days of differentiation. **(D)** Quantification of the number of *MECP2* and *ATRX* foci that overlap with the *XIST* cloud. **(E)** DNA FISH analysis of *MECP2* and *FMR1* in undifferentiated LUHMES cells, combined with RNA FISH against *XIST*. Scale bars are 20 µm in A, 200 µm in C+E.

3.4 Comparison of MeCP2 levels between LUHMES neurons and mouse tissues

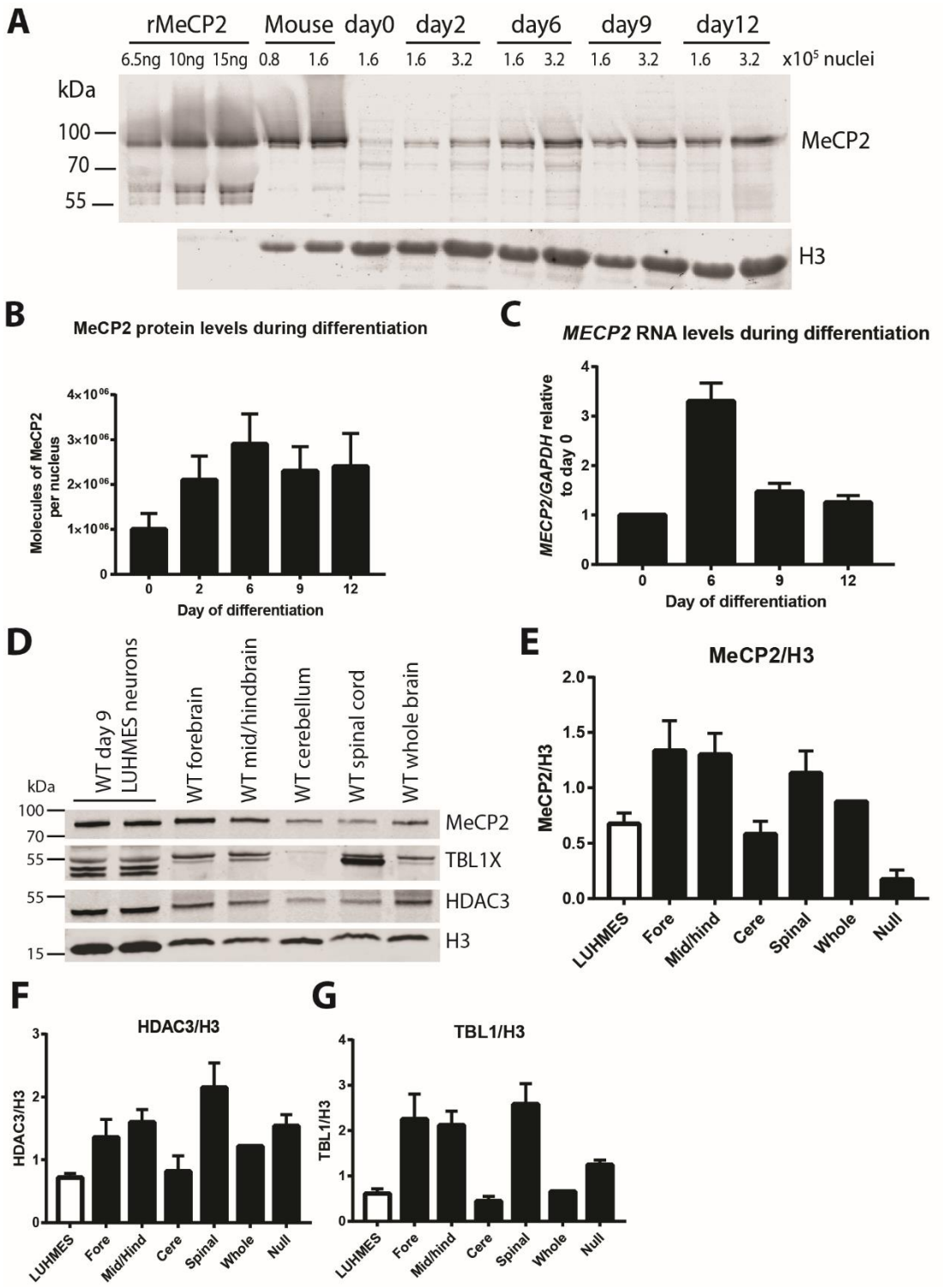
3.4.1 MeCP2 levels during LUHMES cell differentiation

MeCP2 is known to increase its abundance throughout neuronal development and maturation in both mice (Kishi & Macklis, 2004) and humans (Balmer *et al.*, 2003), and I therefore wanted to understand the dynamics of MeCP2 expression during LUHMES cell differentiation. My previous immunofluorescence analysis suggested an increase in the expression level from day 0 to day 12 (Figure 4D). To assess this in a more quantitative manner I performed Western blotting analysis on nuclear extracts from LUHMES cells at different stages of differentiation. By comparing the MeCP2 signal to known loaded amounts of recombinant MeCP2 protein I could estimate the amount of MeCP2/nucleus in LUHMES-derived neurons (Figure 7A). Mouse whole brain nuclear extracts were provided by Dr Sabine Lagger to compare LUHMES MeCP2 levels to that *in vivo* in mice. By comparing the band intensities of MeCP2 in the samples to that of the recombinant protein the nanogram amount of MeCP2 in the loaded sample could be estimated. Then, using Avogadro's number (6×10^{23}) and the molecular weight of MeCP2 (52.3 kDa) the number of MeCP2 molecules per nucleus in each sample was calculated.

As can be seen in Figure 7B, the number of MeCP2 molecules per nucleus increases during LUHMES cell differentiation, reaching a maximum after about 6 days, which coincides with the peak expression of many markers of mature neurons such as Synapsin1, DLG4, βIII-tubulin and NeuN (Scholz *et al.*, 2011). The 3×10^6 molecules of MeCP2/nucleus in 6-day old LUHMES-derived neurons is 5-fold lower than the 15×10^6 molecules/nucleus obtained here for mouse whole brain nuclear extract (Table 9). This number is somewhat higher than a previous estimation by Dr Pete Skene which found that mouse whole brain has 6×10^6 molecules/nucleus (Skene *et al.*, 2010), thus if I compare to the published data LUHMES cells have 50% less MeCP2 protein than the mouse whole brain.

Figure 7. Analysis of MeCP2 protein dynamics during LUHMES cell differentiation.

(A) Example Western blot of a single differentiation of LUHMES cells into neurons, with recombinant MeCP2 and mouse whole brain extract loaded. **(B)** Quantification of molecules of MeCP2/nucleus during the timecourse of LUHMES cell neuronal differentiation. (n=4) **(C)** RT-qPCR analysis of *MECP2* RNA levels during LUHMES cell differentiation. (n=3 as minimum) **(D)** Example Western blot probing for MeCP2, TBL1X and HDAC3 in LUHMES cells and mouse brain regions. **(E-G)** Quantification of ratios from Western blot analysis; MeCP2/H3 (E), HDAC3/H3 (F), TBL1/H3 (G). MeCP2 antibody is M6818 (Sigma) in A+B, or M7443 (Sigma) in D-G. (n=3 as minimum)



There are a number of reasons why the MeCP2 amount in LUHMES-derived neurons could be lower than in mice. Firstly, 9-day old LUMHES-derived neurons are younger compared to a 10-12 week old mouse, and secondly, neurons *in vivo* in the mouse brain are supported by a network of astrocytes and glial cells that aid the maturation of neurons (Johnson *et al.*,

2007; Tang *et al.*, 2013) whereas isolated LUHMES-derived neurons in a dish will not receive that stimulus. In conclusion, the amount of MeCP2 protein in LUHMES-derived neurons increases during differentiation in accordance with *in vivo* development, and reaches a high level of expression that is approximately 5-fold lower than that in mouse whole brain.

Sample	**MeCP2 fg/nucleus**	**MeCP2 molecules/nucleus**
Mouse whole brain	1301.82	1.5×10^7
Day 0 LUHMES	88.12	1.0×10^6
Day 2 LUHMES	184.43	2.1×10^6
Day 6 LUHMES	256.55	2.9×10^6
Day 9 LUHMES	201.26	2.3×10^6
Day 12 LUHMES	209.92	2.4×10^6

Table 9. Semi-quantitative Western blotting to calculate the amount of MeCP2/nucleus in mouse whole brain tissue and during LUHMES cell neuronal differentiation.

Numbers are the average from four separate differentiations which were each loaded on separate gels.

I also assessed the expression dynamics of *MECP2* mRNA during LUHMES cell differentiation and found that, similar to the protein dynamics, the level of *MECP2* mRNA increases during differentiation (Figure 7C). Surprisingly though, the mRNA level drops off after day 6, while the protein amount remains high which perhaps is indicative of a high stability of MeCP2 protein.

3.4.2 MeCP2 protein complex levels in LUHMES-derived neurons

If LUHMES-derived neurons are to be used in the laboratory as a model system for studying MeCP2 and its function in neurons, then it is important to establish that proteins that are known interactors of MeCP2 are expressed. I performed Western blot analysis on HDAC3 and TBL1, two proteins that are part of the NCoR/SMRT repressor complex, which interacts with the NID of MeCP2, and are thought to contribute to MeCP2-mediated gene repression (Gabel *et al.*, 2015; Lyst *et al.*, 2013). For this experiment I obtained regions of mouse brain tissue from Dr Jacky Guy to assess the protein levels *in vivo* as well. Figure 7D shows a representative Western blot, probing for MeCP2, HDAC3 and TBL1, with H3 as the loading control. A first interesting observation is the presence of multiple bands that are picked up by the TBL1 antibody. The X-chromosomal TBL1 protein (also known as TBL1X) has an autosomal paralogue called TBL1R1. The antibody used here is reported on the Abcam website to not cross-react with TBL1R1 yet it was generated against the full-length TBL1 protein. With 86% homology between the two proteins (Choi *et al.*, 2011), perhaps the

multiple bands picked up by Western blot analysis are the two TBL paralogues. While the lower TBL1 band appears to be more abundant in the spinal cord, replicate Western blots using samples from multiple mice established that in all brain regions, the upper band is more abundant (Supplementary Figure 2). There is a third unexplained band in the two LUHMES-derived neuron samples. This could be a third paralogue, a modified version of the protein or non-specific antibody reactivity. Performing a co-IP experiment by immunoprecipitating MeCP2, and probing the extract with the TBL1 antibody might help determine the functional significance of this third band. Of note, the cerebellum contains lower levels of MeCP2 (Figure 7E), consistent with reports from mouse (Ross *et al.*, 2016), as well as low levels of HDAC3 and TBL1 (Figure 7F+G). In conclusion, the two MeCP2-interaction partners HDAC3 and TBL1 are expressed in LUHMES-derived neurons, and the level of these two proteins (as well as MeCP2) is comparable to the mouse brain.

3.5 Development of efficient transfection and single-cell cloning techniques

3.5.1 Testing and optimisation of transfection techniques

A previous publication using LUHMES cells described Lipofectamine-2000 transfection and electroporation of LUHMES cells as being highly variable (Schildknecht *et al.*, 2013). In that study they optimised Nucleofection of cells after differentiation for two days and demonstrated efficient transfection. Their approach is useful for immediate analysis of neurons after Nucleofection, but is not suitable for the generation of a stock of proliferating, genetically modified LUHMES cells, as is the aim of this project. Therefore I tested a number of different transfection and transduction techniques with the proliferative LUHMES cell line.

I began with viral transduction experiments. Lentiviruses have been successfully used to transduce undifferentiated LUHMES cells (Schildknecht et al., 2013) and the technique has been used in our group by Dr Justyna Cholewa-Waclaw. Based on this success I was keen to explore the efficiency of adeno-associated virus (AAV) transduction of LUHMES cells. The benefits of AAVs over lentiviruses are 3-fold. First, as these viruses cannot self-replicate they are safer than lentiviruses and can therefore be produced and used in a Category I, rather than in a Category II tissue culture facility (Samulski & Muzyczka, 2014). Second, as rAAV DNA does not tend to insert into the genome of the host cell, rAAVs do not pose the risk of insertional mutagenesis (Buning et al., 2008). Third, rAAVs are a promising system for developing therapeutic approaches for Rett syndrome as these viruses are effective in

mouse model organisms (Gadalla et al., 2017) and have been used directly in the clinic (Gaudet et al., 2016). As there are a number of different AAV serotypes available, each of which has a different efficacy for different types of cells and tissues, I trialled four separate viruses on LUHMES cells. The scAAV9, ssAAV9 and rAAVrh10 viruses were already present in the lab, and rAAV1/2 was a kind gift from Dr Matt Nolan's lab. These viruses were applied in different titres to LUHMES cells, RNA was acquired after 48 hours, and expression of the virally-provided transgene was assessed by RT-qPCR analysis. Two viruses (ssAAV9 and scAAV9) appeared to be ineffective at delivering the CBA-Mecp2_e1 and MeP-Mecp2_e1 cargo respectively to LUHMES cells (Figure 8A), although detecting the expression of *MECP2* RNA on top of endogenous *MECP2* may have affected the experimental interpretation. However both the rAAVrh10 virus (delivering CAG-Cre) and the rAAV1/2 virus (delivering CAG-EGFP) were effective at infecting and expressing their transgene in LUHMES cells, in a manner that was dose-dependent (Figure 8A). As the rAAV1/2 virus was effective at transducing LUHMES cells, and an efficient published protocol existed for generation of rAAV1/2 viruses (McClure et al., 2011), I used Gateway cloning to generate AAV-compatible plasmids containing EGFP and puromycin resistant (PuroR) cassettes which were both driven by the human PGK promoter (phosphoglycerate kinase). Lipofectamine-2000 mediated transfection of these plasmids into Hek293FT cells confirmed correct cloning of both plasmids (Figure 8B+C), thus I proceeded to make AAV viruses. Both viruses were successful at infecting Hek293FT cells, creating EGFP-expressing Hek293FT cells (Figure 8D) as well as Hek293FT cells that were resistant to puromycin treatment, even after 12 days (Figure 8E). When testing both viruses on

LUHMES progenitor cells neither EGFP expression (

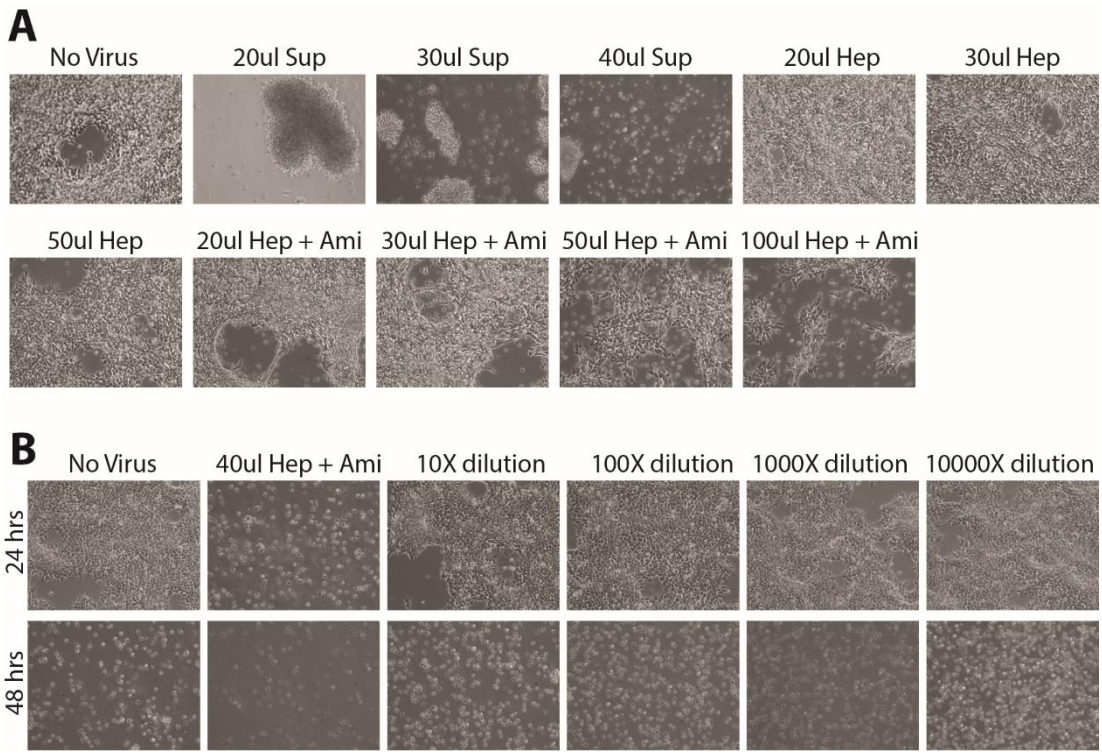


Figure 9A) nor puromycin resistance (

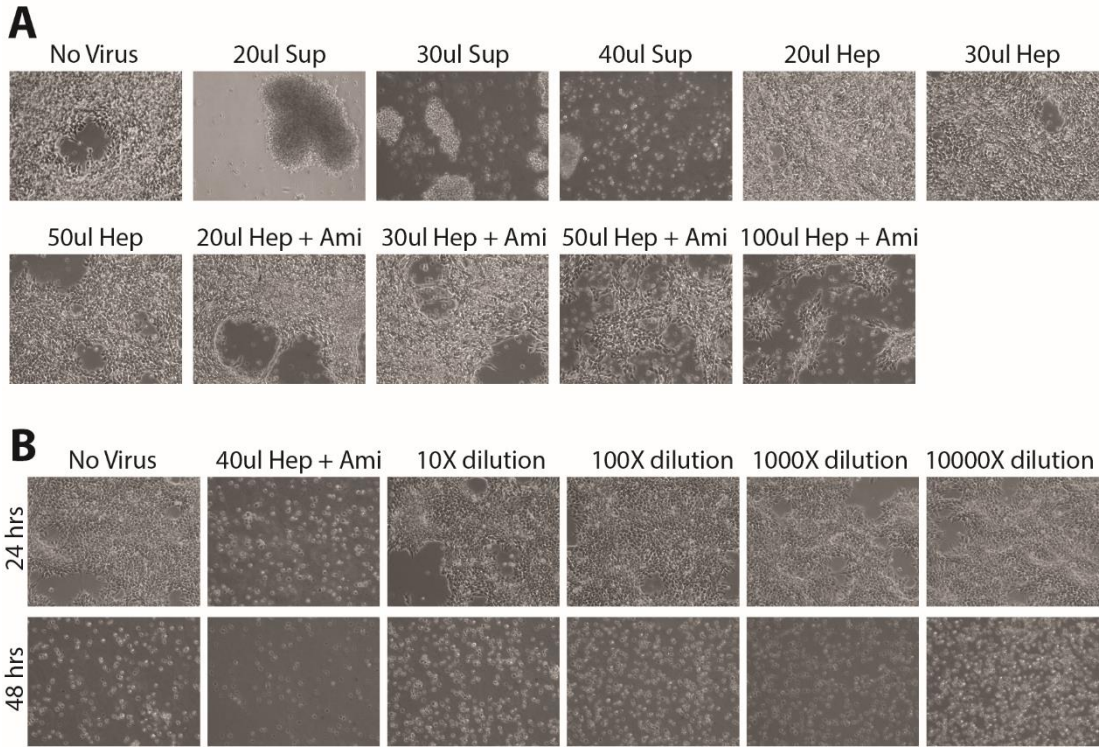
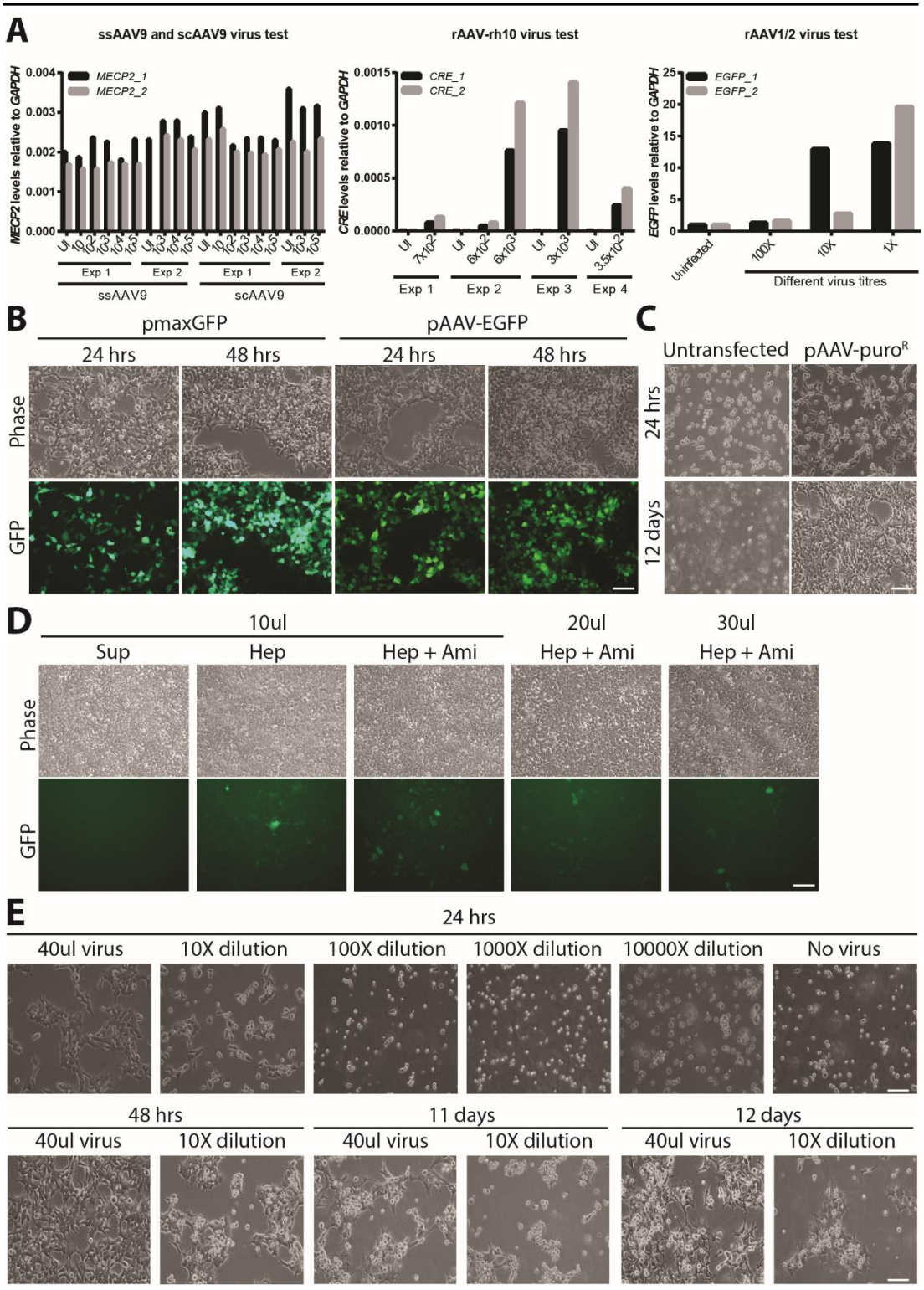


Figure 9B) was achieved, indicating that although functional AAV viruses had been made, neither viral prep effectively transduced LUHMES cells. Perhaps the viral preparations were not pure or concentrated enough to work efficiently on LUHMES cells.

Figure 8. Generation and testing of Adeno-associated viruses in tissue culture.

(A) RT-qPCR analysis of transgene expression from different AAV serotypes after LUHMES infection. UI: uninfected. **(B)** Lipofectamine-2000 transfection of pAAV-EGFP in Hek293FT cells. pmaxGFP is a positive control. **(C)** Lipofectamine-2000 transfection of pAAV-puro^{R} in Hek293FT cells **(D)** AAV1/2-EGFP infection of Hek293FT cells, 48 hours post infection. Sup: AAV immediately harvested from cells. Hep: AAV after purification using a heparin column. Ami: Concentration of AAV viruses on an Amicon spin column. **(E)** AAV1/2-puroR infection of Hek293FT cells using heparin-column filtered and Amicon concentrated virus. Scale bar is 100 µm in all images.


A
ssAAV9 and scAAV9 virus test
rAAV-rh10 virus test
rAAV1/2 virus test
MECP2 levels relative to GAPDH
MECP2_1
MECP2_2
UI
10^2
10^3
10^4
10^5
Exp 1
Exp 2
ssAAV9
scAAV9
CRE levels relative to GAPDH
CRE_1
CRE_2
7x10^2
6x10^2
6x10^3
3x10^3
3.5x10^2
Exp 3
Exp 4
EGFP levels relative to GAPDH
EGFP_1
EGFP_2
Uninfected
100X
10X
1X
Different virus titres
B
pmaxGFP
pAAV-EGFP
24 hrs
48 hrs
Phase
GFP
C
Untransfected
pAAV-puro^R
12 days
D
10ul
20ul
30ul
Sup
Hep
Hep + Ami
E
24 hrs
40ul virus
10X dilution
100X dilution
1000X dilution
10000X dilution
No virus
48 hrs
11 days
12 days

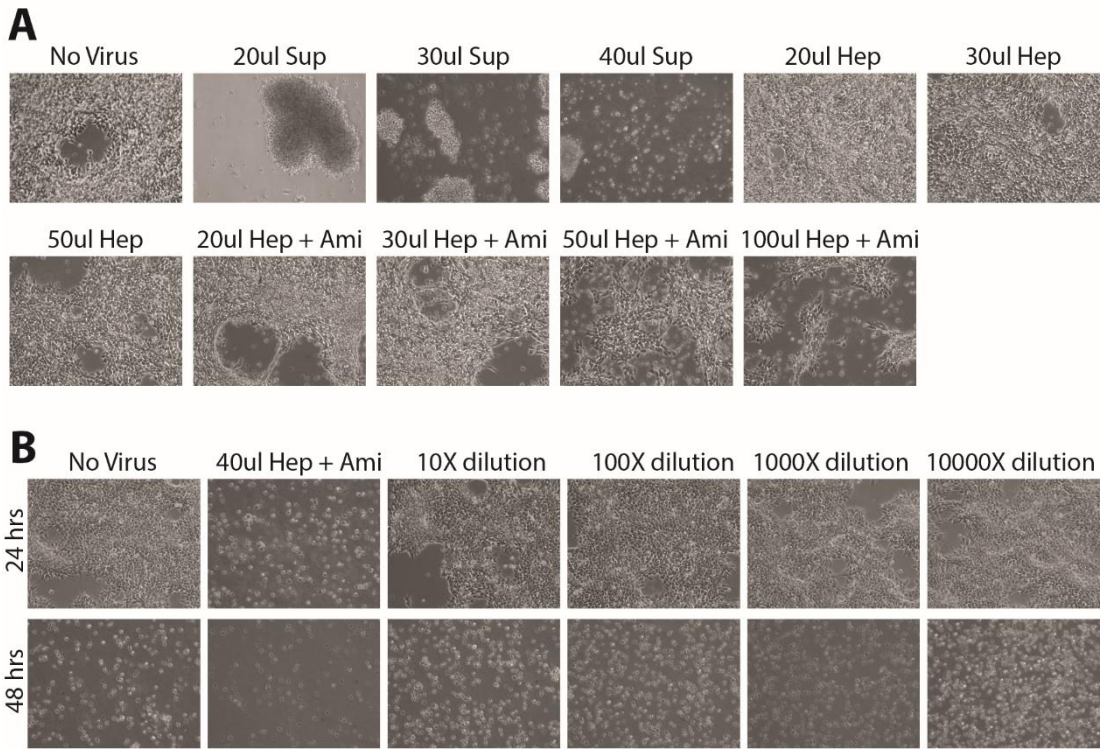


Figure 9. Testing of adeno-associated viruses on LUHMES cells.

(A) AAV1/2-EGFP infection of LUHMES cells. Pictures are taken 48 hours post infection. No GFP fluorescence images are shown as no GFP fluorescence was observed. **(B)** AAV1/2-Puro^R infection of LUHMES cells. Top panel are cells 24 hours post-infection. Bottom panel are cells 48 hours post puromycin selection. Sup: AAV immediately harvested from cells. Hep: AAV after purification using a heparin column. Ami: Concentration of AAV viruses on an Amicon spin column.

As an alternative to AAV, I then optimised Nucleofection of proliferating, progenitor LUHMES cells. Nucleofection is a transfection technique that combines electroporation (the parameters of which are defined by set programs in the Nucleofector device), and resuspension of cells into a buffer that is amenable to transfection. Nucleofection is typically used for cells that are hard to transfect by other means. Five programs were tested using the basic Nucleofection kit for primary neurons in combination with pmaxGFP, the plasmid provided in the kit (Figure legend is over-page.

Figure 10A). RT-qPCR showed extremely high levels of EGFP expression when using program A-33 (Figure legend is over-page.

Figure 10B) and so 18 similar programs were tested to find a program that yielded the best balance of transfection and survival (Figure legend is over-page.

Figure 10C). Notably, LUHMES cells are sensitive to the amount of plasmid that is transfected; 10 µg plasmid induces large amounts of cell death (Figure legend is over-page.

Figure 10D). Furthermore plasmids that are endotoxin-free result in improved cell survival, but phenol-chloroform extraction of the plasmid for further purity did not improve this (Figure legend is over-page.

Figure 10E+F). Finally, it was noticed that incubation of LUHMES cells in a low-calcium media (such as RPMI) for 5 minutes at 37^{o}C post-Nucleofection further improved the survival of LUHMES cells (Figure legend is over-page.

Figure 10F).

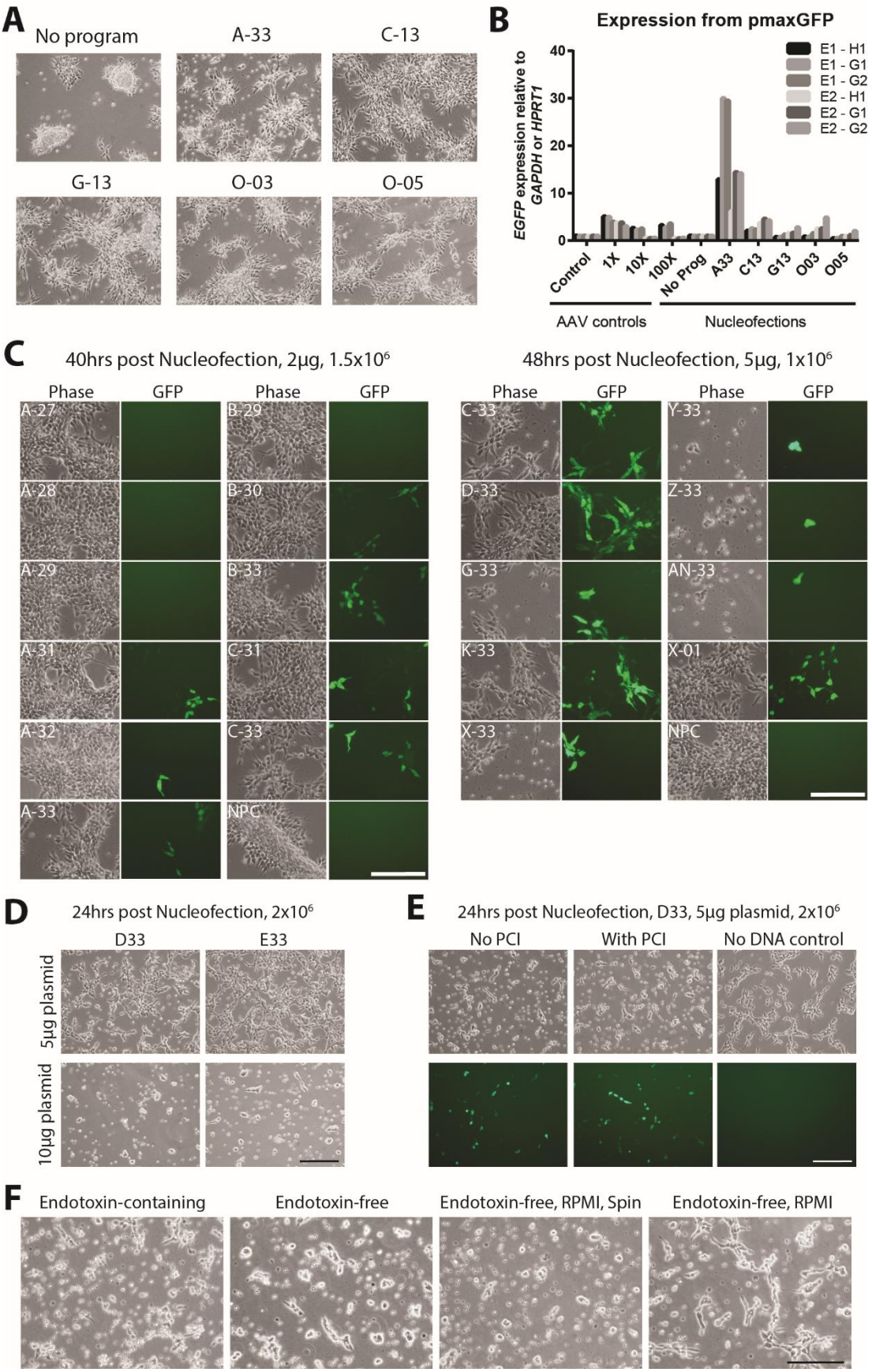


Figure legend is over-page.

Figure 10. Optimisation of Nucleofection in LUHMES cells.

(A) Testing 5 different Nucleofection programs with the Basic Nucleofector kit for primary neurons, as suggested by Lonza. No program = LUHMES cells resuspended in Nucleofection solution but not subjected to an electrical program. **(B)** RT-qPCR analysis of pmaxGFP expression in LUHMES cells 48 hours post-Nucleofection. E1+E2 - two different sets of primers for EGFP. H1 - primer pair for HPRT. G1+G2 - two different sets of primers for GAPDH. **(C)** Trials of various electroporation programs using the Basic Nucleofector kit for primary neurons. **(D)** Comparison of different plasmid amounts on cell viability. **(E)** Comparison of phenol:chloroform:isoamylalcohol plasmid purification on cell viability. **(F)** Comparison of cell viability when endotoxin-free plasmids are used, when the cells are incubated at 37°C in RPMI for 5 minutes (RPMI) and when they are centrifuged before plating (spin). Scale bars are 200 µm.

Other transfection procedures were tested including electroporation (Figure 11A), Neon transfection (Figure 11B), JetPrime polyplus and calcium phosphate transfection but none of these techniques yielded transfected LUHMES cells that were healthy for continued growth. In conclusion, Nuclofection of proliferating LUHMES cells is a reproducible and effective method for transfection of DNA (and potentially RNA) into cells and was considered suitable for the generation of genetically modified cell lines.

3.5.2 Development of FACS sorting for selection of single cells

The LUHMES cell line is a clonal cell line that was established by serial dilution of cells into a 96-well plate. The ability to generate a single clonal cell line containing only the genetically engineered mutation of interest would be extremely beneficial for this project. Therefore, serial dilution into a 96-well plate, colony picking and FACS sorting of single cells into a 96-well plate were each tested. LUHMES cells were not able to form colonies suitable for picking, even with the addition of laminin and B27 to the tissue culture medium to improve colony formation indicating that this strategy was unsuccessful. Serial dilution into a 96-well plate successfully generated single colonies of LUHMES cells that established a clonal cell line, although only 10-15 wells per 96-well plate actually contained colonies and often these were derived from multiple different cells (see Chapter 4.3.2). FACS sorting, however, successfully isolated single LUHMES cells and after being left in culture for 1.5 weeks, individual colonies of LUHMES cells appeared and could be dissociated and transferred to a 24-well plate for expansion.

Thus, the combination of Nucleofection and FACS sorting of single cells into a 96-well plate will be a useful protocol for generating genetically modified, clonal cell lines of interest in this project.

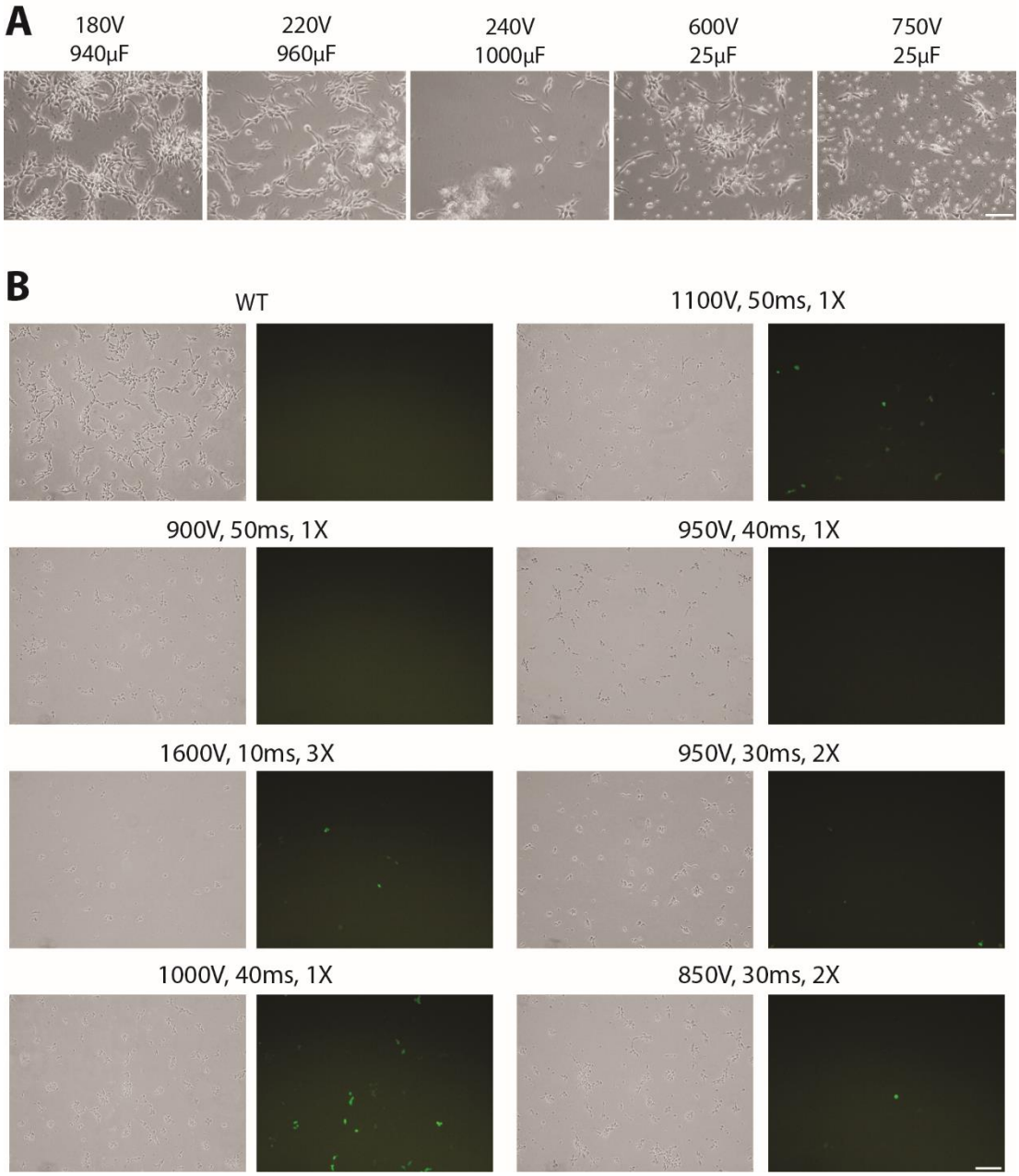


Figure 11. Electroporation and Neon transfection are not suitable methods of transfecting LUHMES cells.

(A) Electroporation of LUHMES cells under different electrical programs. Pictures were taken 40 hours post-electroporation, at this stage no GFP fluorescence was observed. **(B)** Neon transfection of LUHMES cells under different conditions. Picture were taken 24 hours post-transfection. Pulse voltage, pulse width, and the number of pulses are written. Scale bars are 100 µm.

3.6 Exploration of electrophysiological properties of LUHMES-derived neurons

3.6.1 Patch clamping

Several studies using MeCP2-mutant neurons have demonstrated defects in electrophysiological properties. Here I briefly summarise the results from three of these studies in order of increasing phenotypic severity. In 2010, Alysson Muotri and colleagues generated hiPSCs using fibroblasts from four patients containing different Rett syndrome causing mutations, differentiated them into neurons and performed extensive characterisation of these neurons in tissue culture (Marchetto *et al.*, 2010). They performed whole-cell patch clamping experiments using WT and 1155del32 iPSC-neurons, which contain a deletion of 32 nucleotides in the C-terminus of the protein resulting in a frameshift mutation. Both genotypes displayed the same inward sodium and outward potassium currents in response to various voltage depolarisations in voltage-clamp conditions, as well as similar action potentials evoked by current injections in current-clamp conditions. Yet a difference in electrophysiological properties was observed when they assessed spontaneous inhibitory and excitatory postsynaptic currents (siPSCs and sEPSCs). The effects were small, although reduced amplitudes and lower frequencies of sEPSCs were shown in the 1155del32 neurons. In addition to this, Farra and colleagues in 2012 reported reduced sodium and potassium currents and reduced amplitude of sEPSCs in neurons derived from miPSCs from a female heterozygous mouse that has a truncation at position threonine 308 (Farra *et al.*, 2012). Finally, Li and colleagues in 2013 used TALENs to knock-out MeCP2 in hESCs, which they then differentiated into neurons and demonstrated a complete loss of all spontaneous electrical activity using multi-electrode array dishes (Li *et al.*, 2013c).

I was keen to assess the electrophysiological properties of LUHMES-derived neurons, with the hope of being able to establish a reproducible technique that could be used to assess the electrophysiological properties of WT and MeCP2 mutant LUHMES-derived neurons. Whole-cell voltage clamp recordings of LUHMES-derived neurons were previously published demonstrating that they are electrically responsive from day 3, and continue to increase their activity as they mature (Scholz *et al.*, 2011). Furthermore, tests using tetrodotoxin and tetraethylammonium chloride have indicated functioning Na^+ and K^+ voltage-gated channels respectively (Scholz *et al.*, 2011). Here, I collaborated with Dr Sean McKay (Prof Giles Hardingham's lab, Centre for Integrative Physiology, Edinburgh, UK) to perform whole-cell patch clamping experiments. Current-clamp experiments were performed

on 13-day old LUHMES-derived neurons and as can be seen in Figure 12A, neurons were responsive to current injections at this age. As expected, negative current injections (-20pA to -5pA) resulted in hyperpolarisation of neurons, while positive current injections (5pA to 50pA) resulted in depolarisation of neurons. Perhaps if LUHMES cells were co-cultured with astrocytes a more robust electrical activity profile could be induced where the response of the neuron continues to spike over time (Johnson *et al.*, 2007; Kuijlaars *et al.*, 2016; Odawara *et al.*, 2014; Tang *et al.*, 2013). The response gained after only 13 days of differentiation is robust enough, that defects due to MeCP2 mutation could be investigated in future experiments. Unfortunately, due to time constraints, these experiments have not been performed as part of this study.

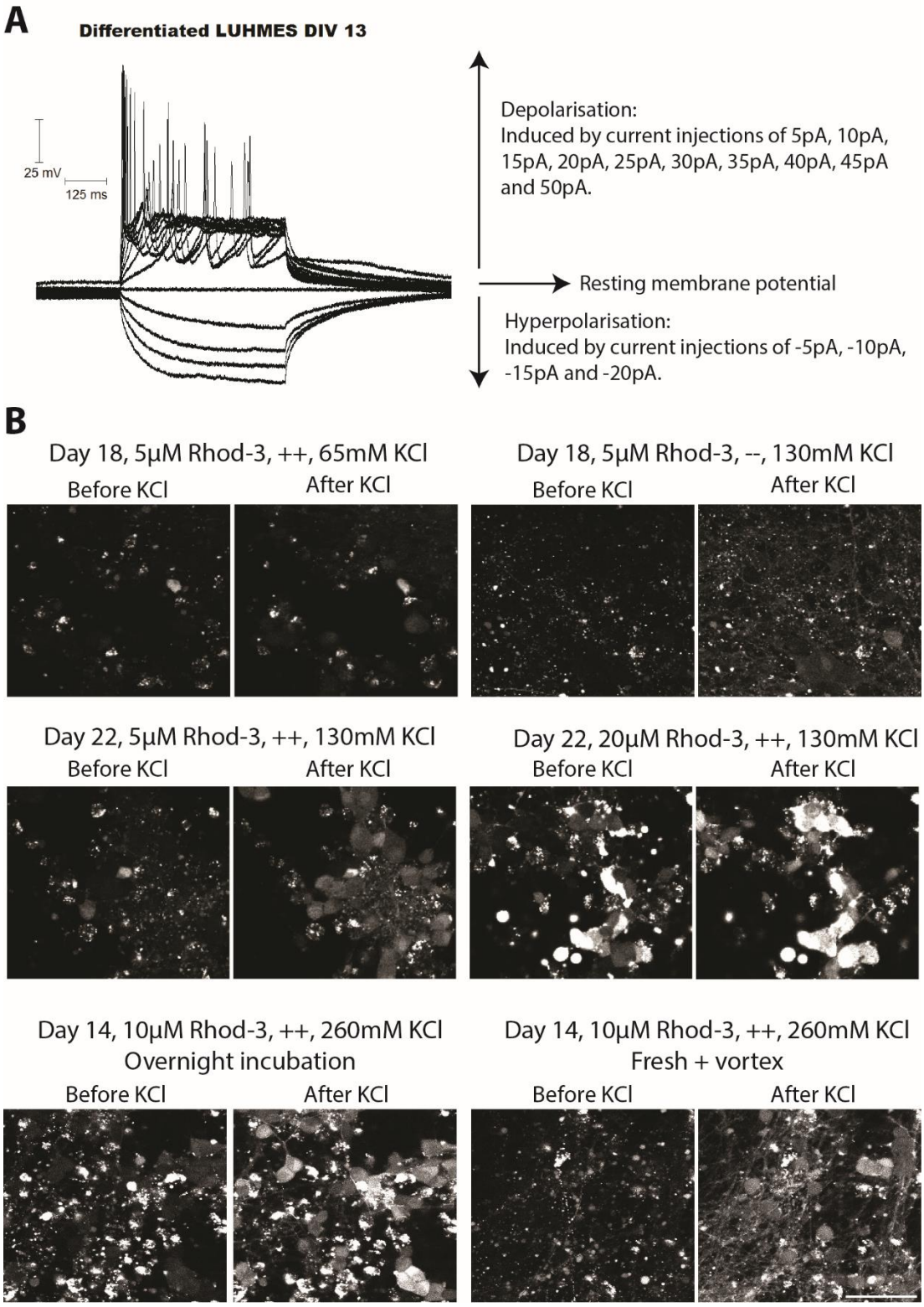
3.6.2 Calcium Imaging

To complement the patch-clamp recordings which focus on a single cell, I also performed calcium imaging experiments in order to assess the dynamics of the whole culture in response to KCl-induced depolarisation. Calcium imaging is a microscopy-based technique that assesses the calcium properties of a cell. For neurons, depolarisation results in the opening of voltage-gated calcium channels, thus allowing extracellular calcium to flood into the cell. This rise in intracellular calcium levels in the neuron triggers pathways that lead to the release of neurotransmitters from the cell, which will bind to receptors on neighbouring neurons and stimulate the depolarisation and formation of action potentials, thus propagating a signal. If the neurons are loaded with a dye that fluoresces upon binding to the calcium ion (Ca^{2+}) then fluorescence microscopy can be used to read the network activity of a population of neurons. Alterations to calcium homeostasis have been reported in some MeCP2 mutant models. Neurons obtained from differentiating hiPSCs from patients with the 1155del32 mutation show a decrease in spontaneous network activity as demonstrated by Fluo-4 calcium imaging (Marchetto *et al.*, 2010), and KCl-induced calcium transients in mouse brain slices evoked larger calcium amplitudes that decayed more slowly in MeCP2-null samples compared to WT (Mironov *et al.*, 2009). Other than using calcium imaging to assess the neuronal properties of MeCP2 mutant LUHMES-derived neurons that will be generated in this study, I was also keen to assess the calcium properties of WT LUHMES-derived neurons as they differentiate and mature.

Figure 12. Stimulus-responses of LUHMES-derived neurons.

(A) Traces of current clamp experiments performed on 13-day old LUHMES-derived WT neurons. **(B)** Trials of Rhod-3 fluorophore for calcium experiments under different conditions. A + or a - sign indicates whether Powerload and Probenecid were included in the Rhod-3 loading mix in that

experiment. "Overnight incubation" indicates that the loading mix was wrapped in foil and rotated overnight at room temperature to encourage thorough mixing. "Fresh + vortex" indicates that that Rhod-3 loading mix was prepared fresh and then vortexed to encourage thorough mixing. Scale bar is 50 µm.



Calcium imaging is not a technique that has been performed in our lab before and there were multiple parameters that needed to be tested to find an optimal protocol; for example calcium indicator of choice, cell loading parameters and imaging parameters. I tested two different calcium fluorophores: Fluo-4 (excitation at 488nm) and Rhod-3 (excitation at 550nm). PowerLoad and probenecid are two supplements that are provided with the Rhod-3 imaging kit; PowerLoad aids the solubilisation of the calcium indicators, while probenecid is an inhibitor of organic-anion transporters that can function to extrude the indicator from cells (Di Virgilio *et al.*, 1990). In addition to testing the two different calcium indicators, I also tested these two supplements for improved imaging.

Three different Rhod-3 concentrations were tested and increases in fluorescence were observed after KCl application. Unfortunately there was high background fluorescence as indicated by speckled dots across the coverslip (Figure 12B). I hypothesized that this background fluorescence could be due to Rhod-3 indicator that was inefficiently dissolved in the HBSS buffer. Unfortunately though, this background fluorescence could not be removed by overnight incubation of Rhod-3 in the HBSS buffer or by vortexing (Figure 12B). The presence of this large amount of background fluorescence would preclude any analysis of these cultures and thus Rhod-3 was exchanged for Fluo-4. Three different Fluo-4 concentrations were loaded into WT LUHMES-derived neurons, all of which produced substantial increases in fluorescence upon KCl addition (Figure 13A, C and E). Addition of 100X PowerLoad and probenecid to the loading buffer did not improve this detection further (Figure 13B, D and F).

As strong fluorescence increases were observed with Fluo-4, I assessed the responsiveness of WT LUHMES-derived neurons to KCl during differentiation. Undifferentiated and 4-day old neurons were not responsive, but neurons that were 10 days old and above produced strong calcium transients (
Figure 14). As 3-day old neurons are reported to be electrically active as judged by patch clamp experiments (Scholz *et al.*, 2011), it is possible that at this early age LUHMES-derived neurons are electrically active and can produce action potentials, yet they do not express the complement of voltage-gated calcium channels required for calcium imaging experiments. Alternatively, the Fluo-4 indicator might not be loaded at a high enough concentration to detect small changes in calcium transients in these cells. Treatment of 4-day old neurons with a calcium ionophore (A-23187) to increase the permeability of the cellular membrane to divalent cations would allow Ca^{2+} to enter the neurons in the absence of

voltage-gated calcium channels. If fluorescence increases were observed, it could be concluded that 4-day old LUHMES neurons lack the required calcium channels.

In order to perform quantitative calcium imaging, which would allow meaningful comparisons between mutant and WT samples in the future, I attempted to establish calcium calibration curves for 2 µM Fluo-4 in 21-day old WT LUHMES-derived neurons. This involved adding buffers containing different concentrations of Ca^{2+} to cells in the presence of the A-23187 calcium ionophore to allow the extra-cellular applied Ca^{2+} to enter the cell in the absence of KCl-induced depolarisation. As a result the amount of fluorescence from the Fluo-4 dye will be directly proportional to the amount of Ca^{2+} that was present in the extra-cellular HBSS buffer. Initially, fluorescence increases were observed when using the 0 mM Ca^{2+} buffer (Figure 15A), suggesting that intracellular stores of calcium were being released by the calcium ionophore and were binding to Fluo-4. By first depleting all intracellular stores by incubation with 2 µl A-23187 in 400 µl 0 mM Ca^{2+} for 500 frames, and then adding the required calcium concentration buffer with A-23187, a gradient of fluorescence changes could then be obtained with increasing Ca^{2+} concentrations (Figure 15B+C). Quantification of this data for two replicates produced two different calcium calibration curves (
Figure 16). The first set of data produced a calibration curve (
Figure 16A), that in the log scale forms a straight line (
Figure 16B), with only one data point (0.5 mM Ca^{2+}) deviating from the trend line. The image for this particular calcium concentration (Figure 15Biv) shows a lack of fluorescence thus explaining the deviation from the trend line. The lack of fluorescence could be due to improper focussing of the sample prior to imaging. The 2nd set of data does not form a linear trend line with increasing calcium concentration (
Figure 16D). In order to investigate the reason for the scattered data in this experiment, the resting fluorescence level of individual cells post intracellular store depletion was plotted (
Figure 16E). This showed that the resting fluorescence value was more variable among and between coverslips in the 2nd experiment compared to the 1st, and this probably accounts for the lack of a reliable calcium calibration curve.

Unfortunately, due to time restraints, replicate calcium calibration curves were not generated and calcium imaging of WT and mutant neurons was not performed in this project. To conclude, calcium imaging has been optimised for use in LUHMES-derived neurons and it has revealed an age-dependent increase in the KCl-induced calcium response of WT neurons.

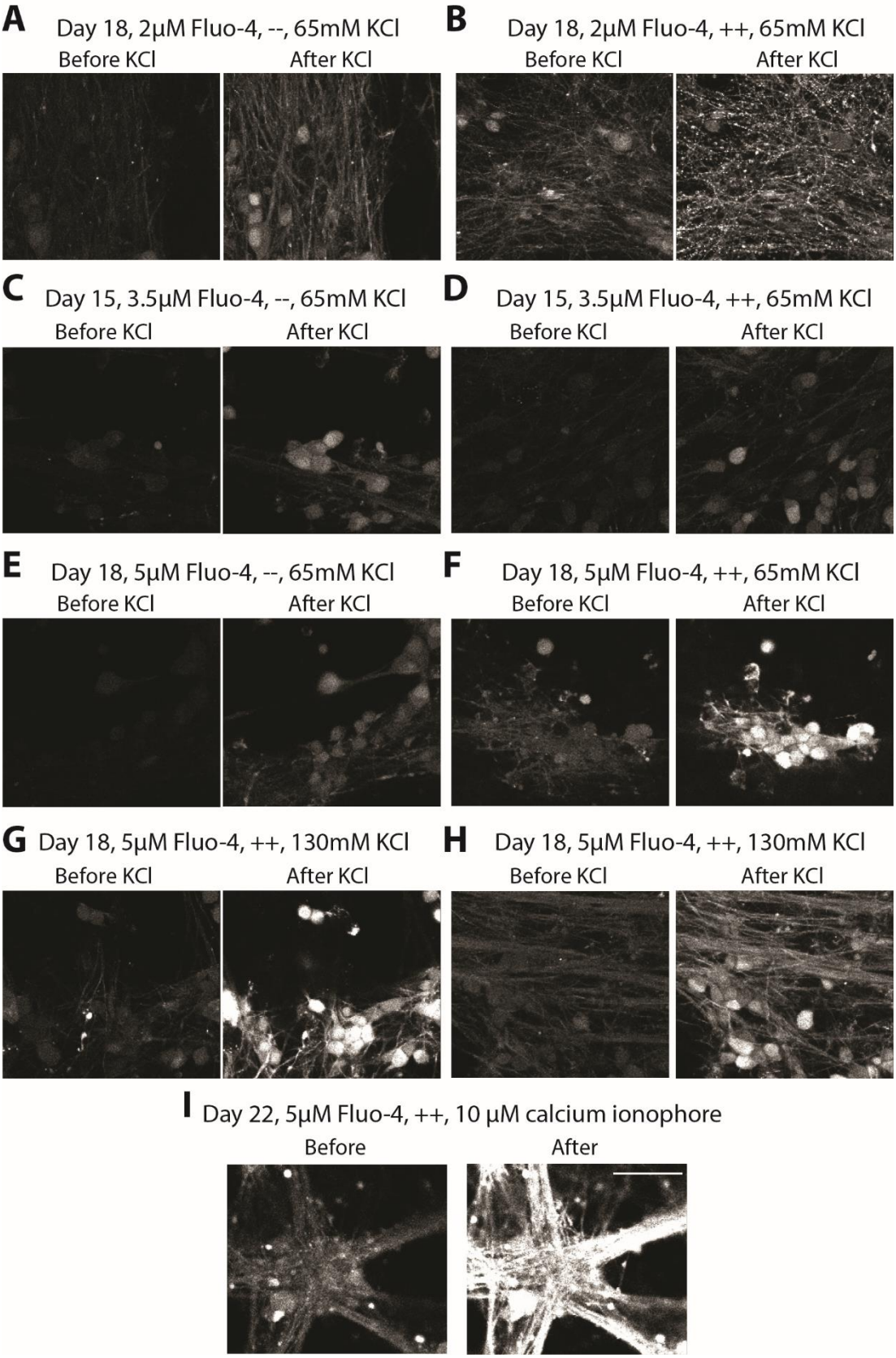


Figure 13. Optimisation of Fluo-4 for calcium imaging.

(A-H) Trials with Fluo-4 fluorophore under different conditions. **(I)** Use of calcium ionophore is a positive control. Scale bar is 50 µm.

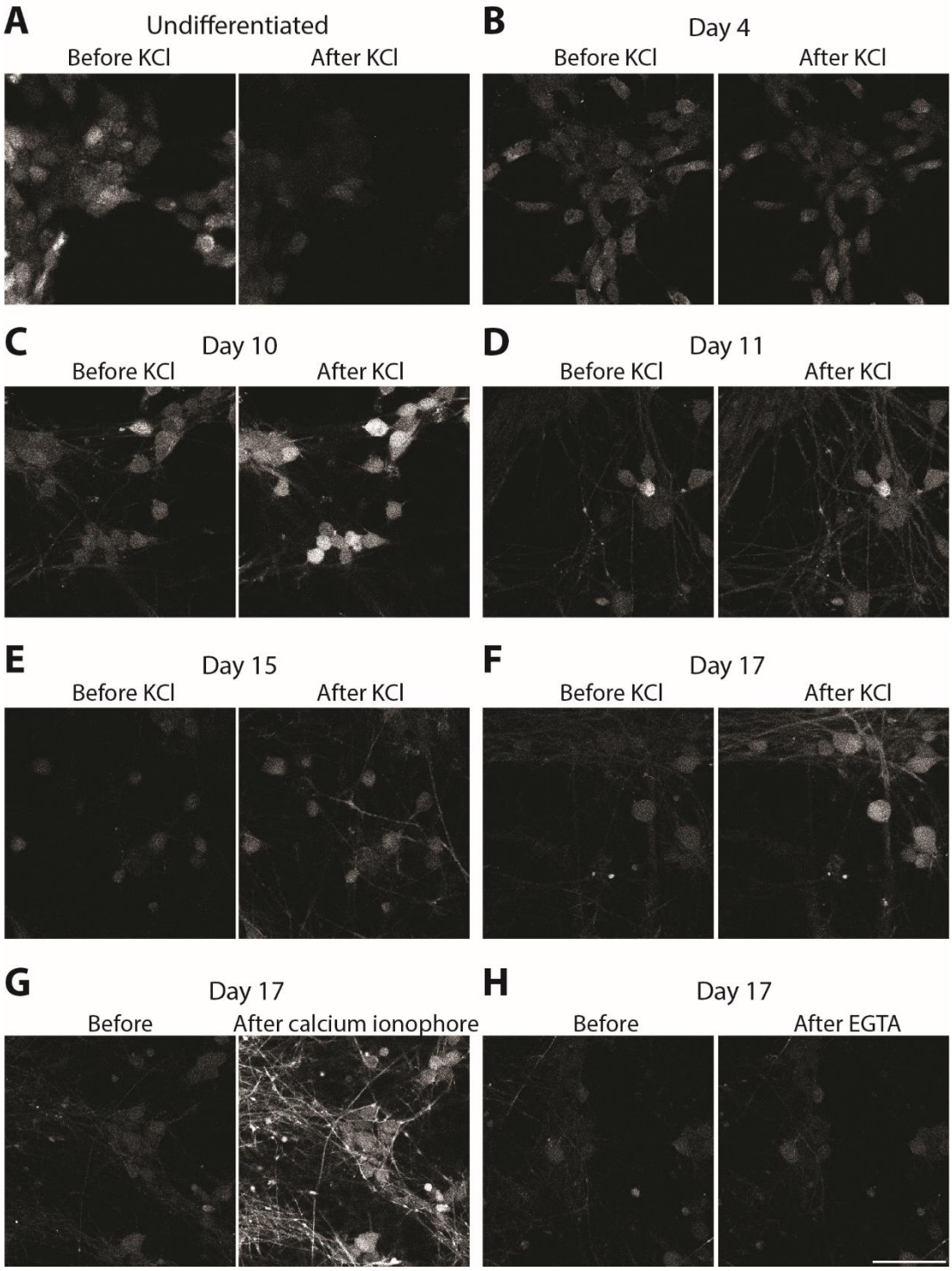


Figure 14. KCl-induced calcium transients increase in LUHMES-derived neurons during differentiation.

Experiments using 2 µM Fluo-4 without additives at different stages of LUHMES cell neuronal differentiation. Calcium ionophore is a positive control and EGTA is the negative control. Scale bar is 50 µm.

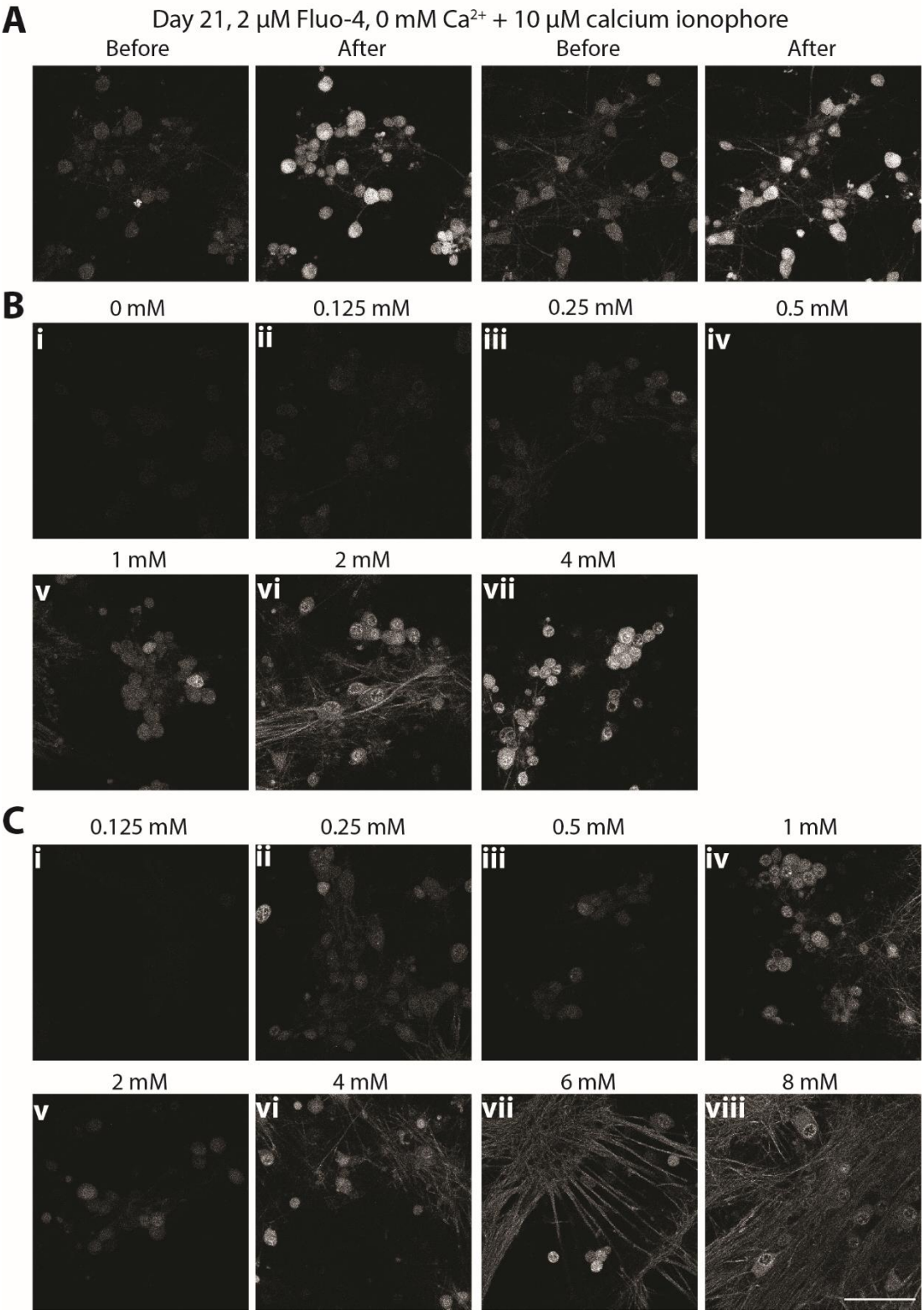


Figure 15. Calcium calibration images of Fluo-4 in 21-day old LUHMES-derived neurons.

(A) Initial tests using calcium ionophore in a 0 mM Ca^{2+} buffer. Images from two coverslips are shown. **(B+C)** Images from two calcium calibration experiments. Neurons were first depleted of intracellular calcium ion stores. Concentrations above the picture indicate the concentration of Ca^{2+} that was present in the HBSS buffer. Scale bar is 50 µm.

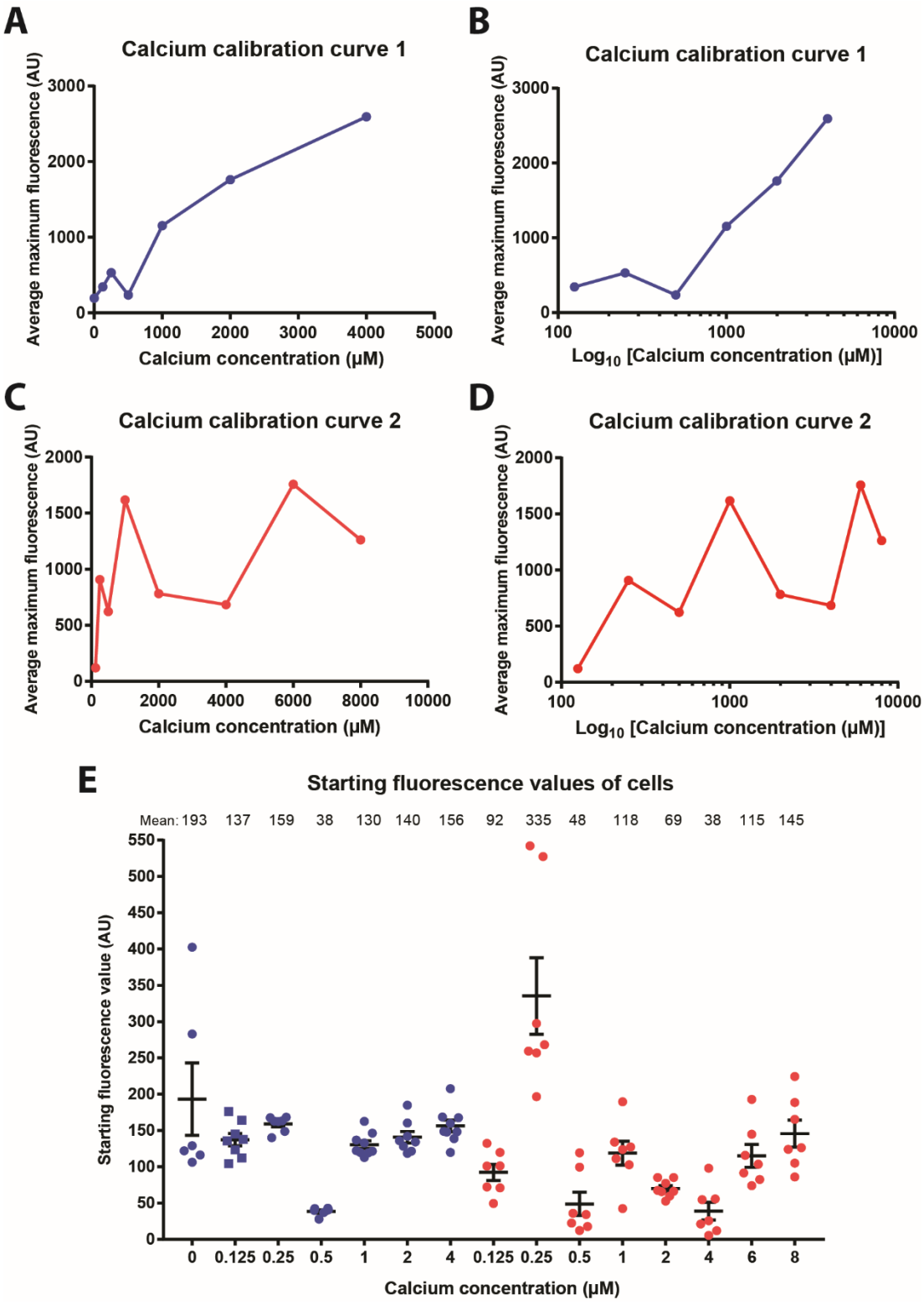


Figure 16. Unreliable calcium calibration curve is due to variable baseline fluorescence levels in LUHMES-derived neurons.

(A+B) Results from first set of calcium imaging experiments **(C+D)** Results from second set of calcium imaging experiments. **(E)** Starting fluorescence values of cells after depletion of intracellular Ca2+ stores. Blue samples are those from the 1st experiment (corresponding to A+B),

red samples are those from the 2nd experiment (corresponding to C+D). Each datapoint is a single cell, mean and SEM are shown as error bars, mean values are indicated above the graphs.

3.7 Discussion

This chapter has dealt with the characterisation of LUHMES cells and their derived neurons as a model system for studying MeCP2 and Rett syndrome. No hypothesis was being tested here, yet the insights gained in this chapter are very promising for the use of LUHMES cells as a neuronal tissue culture system and have laid the groundwork for the future use of this cell line by us and others. One drawback for many tissue culture cell lines is the fact that they are generated from cancerous tissue and therefore have an unstable and highly duplicated genome. This is the case for workhorse cell lines such as HeLa and Hek293T cells, however is also the case for some widely used neuronal cell lines such as SH-SY5Y (Krishna *et al.*, 2014). The demonstration here that LUHMES cells have a normal karyotype is encouraging from multiple perspectives; not only does it make future genetic targeting experiments less complicated, it also establishes these cells as a realistic version of the *in vivo* neurons that they represent. In order to further assess the genomic integrity of LUHMES cells, SNP arrays to detect microdeletions and other genomic rearrangements could be performed (Nowak *et al.*, 2009). In particular analysis of young passage number LUHMES cells (for example P8, which would correspond to the RNA FISH data) as well as old passage number LUHMES cells (for example P25) would give confidence in the genomic stability of the cells as they are kept in tissue culture.

The generation of an efficient and reproducible transfection technique will prove to be vital for the genetic manipulation of LUHMES cells by CRISPR technologies. Furthermore, Nucleofected LUHMES cells are amenable to FACS sorting of single-cells into a 96-well plate two days post-Nucleofection and from here multiple clonal cell lines can be established.

Finally, the demonstration that techniques such as calcium imaging and patch clamp experiments can be performed with LUHMES-derived neurons establishes that this cell line can be used to explore the electrophysiological aspects of mutant neurons. Such experiments would complement *in vivo* datasets in order to bridge the gap between disease-causing mutations, neuronal malfunction and whole organism pathophysiology.

Chapter 4

4. Efficient and versatile CRISPR engineering of human neurons in culture

4.1 Introduction

Since demonstrating efficient transfection and single-cell cloning of LUHMES cells, the next aim of this project was to genetically manipulate the endogenous *MECP2* locus using CRISPR technologies, in order to create a panel of *MECP2* mutant cell lines. The ability to knock-in (KI) RTT-causing mutations into the LUHMES cell line will generate useful models that can be studied for biochemical, transcriptional and neurological phenotypes. Such studies will help to gain insight into the function of MeCP2 in human neurons, and the molecular consequences that lead to Rett syndrome when this protein is mutated.

Part of this work was published in Wellcome Open Research in November 2016 (specifically Chapter 4.3) and therefore I have gained permission from my supervisor, Prof Adrian Bird, as well as the co-authors to include the results from the paper as part of this thesis. There are, however, some experiments in the paper that I did not perform and therefore other author's contributions are detailed in the appropriate figure legends throughout this chapter.

4.2 Testing of parameters in Hek293FT cells

In order to first trial sgRNAs and to confirm correct sgRNA design, I tested the CRISPR system in Hek293FT (human embryonic kidney) cells which are an easy to transfect cell line with high transfection efficiencies (~80%) (Lagger *et al.*, 2017). Seven sgRNAs targeting exon 4 of the *MECP2* locus, which lies on the X chromosome, were designed and individually cloned into a plasmid, pX330, that also expresses Cas9 (Figure 17A+B). Hek293FT cells were transfected and a T7 Endonuclease I (T7E1) assay was performed

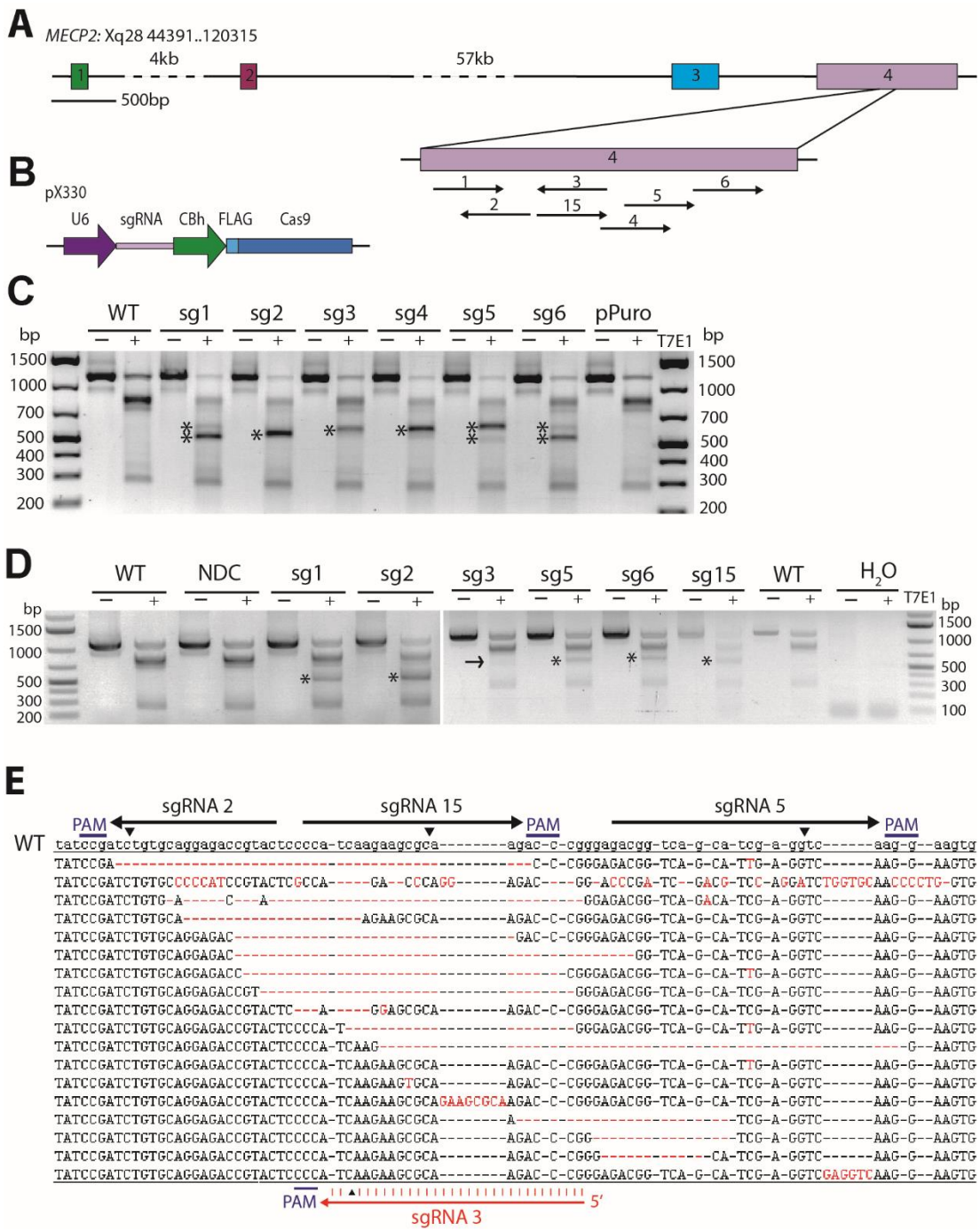


Figure 17. The endogenous *MECP2* locus in human LUHMES cells can be modified by the CRISPR/Cas9 gene editing system.

(A) Schematic representation of the *MECP2* locus with seven sgRNAs targeting both strands within exon 4. **(B)** Schematic representation of the plasmid used to express Cas9 and sgRNA constructs. **(C)** T7E1 assay of sgRNAs targeting *MECP2* in human Hek293FT cells. WT: wild-type cells that have not been transfected. pPuro: cells that have been transfected with a plasmid conferring puromycin resistance, but not a CRISPR plasmid. Bands specific to samples transfected with Cas9 and sgRNA are marked with an asterix. **(D)** T7E1 assay of sgRNAs targeting *MECP2* in human LUHMES cells. WT: wild-type cells that have not been transfected. NDC: cells that have undergone Nucleofection without the addition of plasmids. Bands specific to samples transfected with Cas9 and sgRNA are

marked with an asterix. **(E)** Sequencing of the MECP2 locus post-CRISPR transfection in LUHMES cells. Genomic DNA from wild-type cells is shown in lower case at the top. PAM sites are indicated with a blue line. Arrowheads indicate sites of double-strand break for each sgRNA. Sequencing also confirms the presence of the sgRNA 3 targeting sequence in LUHMES gDNA.

which confirmed correct targeting of all sgRNAs to the *MECP2* locus (Figure 17C). LUHMES cells were then Nucleofected with the same plasmids, positively transfected cells were selected by an overnight puromycin selection (with resistance conferred by co-transfection of a 2nd plasmid pAAV-Puro^{R}), and genomic DNA was harvested 3-5 days post Nucleofection. The T7E1 assay was performed to asses for INDEL formation at the *MECP2* locus (Figure 17D), and genetic alterations were confirmed by sequencing of genomic DNA (Figure 17E). As can be seen, sgRNAs 1, 2, 5, 6 and 15 all induce INDEL formation in LUHMES cells with a good efficiency. On the other hand, sgRNA 3 does not appear to have induced INDEL formation at a high enough frequency for the T7E1 assay to reasonably detect (arrow in Figure 17D). The presence of the sgRNA 3 targeting sequence was confirmed in LUHMES cells by genomic DNA sequencing (Figure 17E), indicating that the low INDEL efficiency is not due to SNPs. These results indicate that sgRNA efficiency can vary quite dramatically between two alternative human cell lines (also reported in (Swiech *et al.*, 2015)), and reinforces the conclusion that multiple sgRNAs should be designed for a single targeting experiment, and they should be tested for functionality in the cell line of interest.

4.3 Efficient and versatile CRISPR engineering of human neurons in culture to model neurological disorders

4.3.1 Introduction

The advent of technologies that introduce targeted mutations into the genome has dramatically changed the way in which genetic diseases can be modelled and studied. The most recent development in the genome editing field, the clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR associated 9 (Cas9) system, has proven to be extremely successful, due in part to its ease of use and efficient implementation in a variety of cell lines (Cong *et al.*, 2013; Jinek *et al.*, 2013; Liang *et al.*, 2015a) and model organisms (Friedland *et al.*, 2013; Gratz *et al.*, 2013; Hai *et al.*, 2014; Hwang *et al.*, 2013; Jiang *et al.*, 2013b; Li *et al.*, 2013b; Niu *et al.*, 2014b; Wang *et al.*, 2013a). The coupling of CRISPR gene editing technology with human induced pluripotent stem cells (iPSCs) has rapidly expanded the number of neurological disorders that can be modelled in a human

neuronal background and is allowing researchers to probe the underlying molecular mechanisms in unprecedented detail (Howden *et al.*, 2016; Merkert & Martin, 2016; Mungenast *et al.*, 2016). In particular, the ability to genetically modify a hiPSC line to create isogenic cell lines, which are genetically identical (bar the disease causing mutation), and differentiate these into neurons for phenotypic analysis is extremely powerful. However, despite advances in hiPSC culture and neuronal differentiation protocols, there are still some limitations to this strategy. One obstacle is the large variability of clonal iPS cell lines when they are derived, which can have negative downstream effects on CRISPR targeting efficiency, single cell cloning and particularly on phenotypic outcomes. Furthermore, there is still debate as to the robustness of the epigenome in iPSCs after reprogramming (Kim *et al.*, 2010; Ohi *et al.*, 2011).

Alternative human neuronal progenitor cell lines are available including the SH-SY5Y line and neural stem cells derived from fetal human brain or human embryonic stem cells. Yet each of these models has drawbacks. SH-5YSY cells are a neuroblastoma cell line with multiple chromosomal duplications and deletions (Krishna *et al.*, 2014) and neural stem cells take a long time to mature during the differentiation process, expressing markers specific for neuronal progenitors for at least four weeks (Shin & Vemuri, 2010; Sun *et al.*, 2008; Tong *et al.*, 2016). The LUHMES neuronal progenitor cell line is a recent alternative that is proving to be highly useful in the neuroscience field (Lotharius *et al.*, 2002). These female "pre-neuronal" cells are forced to proliferate in an immature state by expression of the retroviral element v-myc (Hoshimaru *et al.*, 1996). V-myc expression is under the control of tetracycline so by simple administration of the drug to cell culture medium, LUHMES cells undergo a rapid and robust differentiation into a homogeneous population of electrically active, post-mitotic, mature dopaminergic neurons in just 1 week (Lotharius *et al.*, 2002). . The resulting neurons have thus far been used to model Parkinson's disease (Lotharius *et al.*, 2005; Xiang *et al.*, 2013), for cytotoxicity assays (Tong *et al.*, 2016) and for technology development (Dinh *et al.*, 2013; Hughes *et al.*, 2014; Ilieva *et al.*, 2013).

In order to make this cell line more widely applicable for the neuroscience field, it would be beneficial to routinely genetically modify LUHMES cells to create a variety of cell lines for disease modelling and drug-screening purposes. Historically genetic manipulation via homology-directed repair (HDR) of somatic cells has been difficult, with the most successful approaches involving rAAV-delivered homology arms to produce targeting efficiencies of ~1% (Porteus & Baltimore, 2003; Porteus *et al.*, 2003; Russel & Hirata, 1998). With the

advent of CRISPR technologies, HDR targeting efficiencies in somatic cells has increased somewhat, although to different extents in different systems, for example 1.3% in primary neonatal fibroblasts (Lin *et al.*, 2014), 1.8% *in vivo* by AAV delivery to mouse lung tissue (Platt *et al.*, 2014) and 17% in T cells using Cas9 ribonucleoprotein complexes (Schumann *et al.*, 2015).

Here we describe a robust and reproducible method for the efficient transfection of LUHMES cells, and demonstrate various ways in which this cell line can be genetically manipulated using CRISPR engineering to create human models for the study of neurological disorders.

4.3.2 Results

Transfection of proliferating LUHMES cells using Nucleofection

In this study we sought to edit the endogenous genome of the pre-neuronal somatic LUHMES cell line in three ways: i) by disrupting a target gene; ii) by introducing discrete mutations into the protein coding region; iii) by adding a relatively large protein tag to generate a fusion protein (Figure 18A). Karyotyping confirmed that LUHMES cells have a normal diploid chromosome complement (Supplementary Figure 3A) and RNA FISH demonstrated X inactivation to be established in the pre-neuronal cells prior to differentiation (Supplementary Figure 3B). As a first step in editing the genome, a reproducible method of plasmid transfection needed to be established for these cells. LUHMES cells have proven to be difficult to transfect (Schildknecht *et al.*, 2013) and as a result previous studies relied on lentiviruses. In our hands transient transfection methods such as electroporation, Lipofectamine-2000, Neon transfection and JetPrime Polyplus all resulted in cell death or extremely low levels of transfection (unpublished report, RRS, JCW, AB). Nucleofection has previously proven to be successful, but only after differentiating the cells for 2 days prior to trypsinisation and Nucleofection (Schildknecht *et al.*, 2013). For the generation of genetically modified cell lines, this protocol is undesirable as any transfected cells will be immediately differentiated into post-mitotic neurons, and no stock of proliferating, genetically modified cells remains. In order to optimise Nucleofection conditions in proliferating LUHMES neuronal progenitor cells, the Amaxa Basic Nucleofector Kit for primary neurons was used with 20 different Nucleofection programs to find the optimal balance between transfection efficiency and cell viability (Supplementary Figure 3C). Program D33 reproducibly yielded transfection efficiencies of 25-30%, as judged by the number of GFP-positive cells in the population (Figure 18B). Of note,

transfection of plasmids that have been prepared in an endotoxin-free environment resulted in increased cell viability, but purification of plasmids by ethanol precipitation did not improve this (Supplementary Figure 3E+F). In this way we achieved efficient and reproducible transfection of proliferating LUHMES cells using exogenous plasmids.

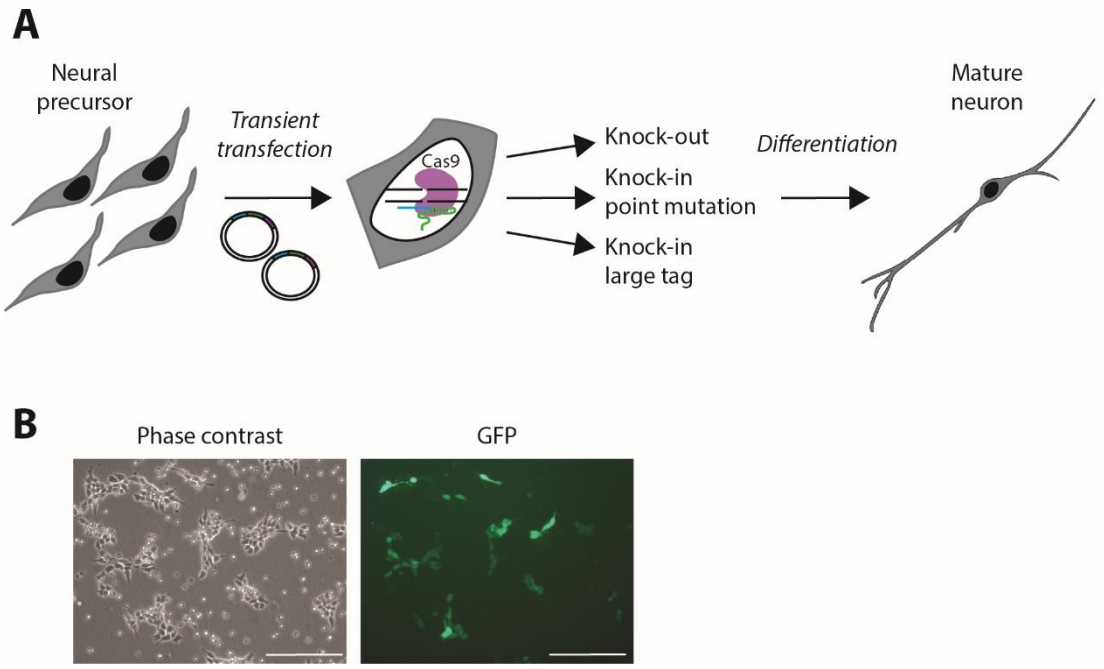


Figure 18. Nucleofection for efficient transfection of LUHMES cells.

(A) The aims of this study are to develop methods for genetic manipulation of the LUHMES pre-neuronal cell line, which in itself requires methods of transient transfection to be developed. After successful genetic modifications have been confirmed, LUHMES cells can then be differentiated into mature neurons for study. **(B)** Representative images of plasmid transfection of LUHMES cells using program D33 with a Nucleofector device. Pictures were taken 47 hours post-Nucleofection. Scale bar is 200 µm.

Generation of a MeCP2 knock-out cell line

We next tested the ability of CRISPR/Cas9 to generate a knock-out LUHMES cell line. For this the *MECP2* locus was chosen. The MeCP2 protein is highly expressed in neurons (Shahbazian *et al.*, 2002a; Skene *et al.*, 2010) and mutations within this protein lead to the autism-spectrum disorder Rett syndrome (Amir *et al.*, 1999). Multiple mouse models of Rett syndrome have been developed, including mice containing Rett syndrome-causing point mutations (Brown *et al.*, 2016) as well as knock-out alleles (Chen *et al.*, 2001; Guy *et al.*, 2001). The *MECP2* gene has four exons, with different isoforms being expressed from exons 1 and 2. As exon 3 is the first shared exon among all isoforms, this was chosen for targeting in order to ablate all MeCP2 protein isoforms. Two sgRNAs were designed within exon 3

(Figure 19A) and were individually cloned into a plasmid that also encodes Cas9 and a puromycin resistance gene (Figure 19B) (Sanjana *et al.*, 2014). LUHMES cells were Nucleofected (Supplementary Figure 4A) and after selecting for positively transfected cells using puromycin resistance both sgRNAs were confirmed to be functional by the T7E1 assay (Figure 19C) and single-cell colonies were established by serial dilution in 96-well plates. The genomic DNA from single cell colonies was extracted and sequenced in order to identify potential positive KO clones. The genomic DNA sequencing from two different cell lines is shown in Figure 19D, KO1 has a homozygous deletion of 9bp whereas KO2 has a heterozygous deletion of 14bp, with the second allele being unaltered. As *MECP2* resides on the X chromosome and LUHMES cells are female cells with one X chromosome already in the inactive state (Supplementary Figure 3B), the homozygous 9bp deletion in KO1 suggests that the inactive X chromosome can be edited by the CRISPR/Cas9 system. Overall, out of 13 colonies that were sequenced, 11 contained INDELs thus giving a targeting efficiency of 85%.

To determine the genotype of the actively expressed *MECP2* mRNA in these cell lines, cDNA sequencing was performed (Figure 19E). The 14bp deletion allele in KO2 appears to reside on the active X chromosome as all cDNA sequence reads from this cell line contained this out-of-frame deletion, highly indicative of a protein KO phenotype. Surprisingly the 9bp in-frame deletion in the middle of exon 3 of KO1 resulted in the whole of exon 3 being removed from the mature mRNA transcript, causing exons 2 and 4 to be spliced together in-frame. Western blot analysis confirmed the complete absence of any full length MeCP2 protein in both cell lines (Figure 19G). In order to identify clones that might contain truncated protein, Western blot analysis was performed using two different antibodies, one against the N-terminus of MeCP2 and another against the C-terminus, and this revealed that KO1 has very low levels of a truncated protein (Supplementary Figure 4B). Even though this cell line cannot technically be referred to as a protein KO cell line, the extremely low MeCP2 protein level that remains (and the removal of critical residues via deletion of exon 3) probably results in a cell line that is phenotypically null, as has been observed in mice (Chen *et al.*, 2001). Finally, the top off-target loci for each sgRNA were sequenced for off-target INDEL formation and as expected based on recent findings in hiPSCs (Paquet *et al.*, 2016), no off-target cutting was observed (Figure 19H). These experiments confirm Cas9-induced INDEL formation to be successful, specific and highly efficient in LUHMES cells.

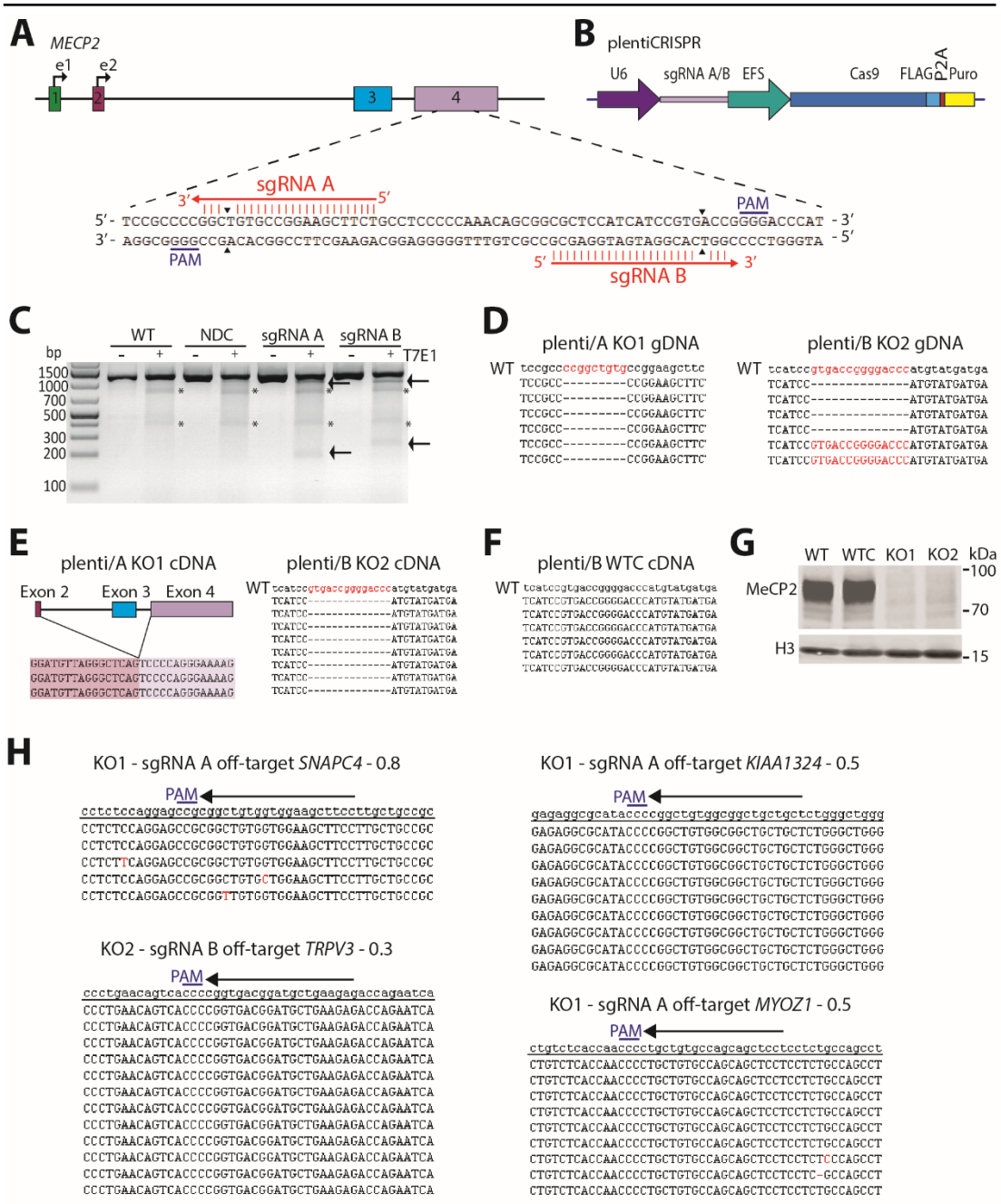


Figure 19. Generation of MeCP2 knock-out LUHMES cell lines.

(A) Schematic representation of the MECP2 locus with the targeting sequence of sgRNA A and sgRNA B labelled. Arrowheads indicate sites of double-strand break. **(B)** Schematic representation of the plasmid used to deliver Cas9 and sgRNA to LUHMES cells. **(C)** T7E1 assay of sgRNA A and sgRNA B. WT: non-transfected wild-type cells. NDC: wild-type cells Nucleofected without any DNA. Asterisks indicate non-specific bands. Arrows indicate bands specific to samples containing Cas9 and sgRNA. **(D)** Sequencing of genomic DNA from two LUHMES single-cell clones. Wild-type genomic DNA is in lower case at the top. **(E)** Sequencing of cDNA from two LUHMES single-cell clones. Wild-type cDNA is in lower case at the top. **(F)** Sequencing of genomic DNA from a single LUHMES unedited clone that was transfected with Cas9 and sgRNA B. Wild-type gDNA is in lower case at the top. **(G)** Western blot performed by Dr Justyna Cholewa-Waclaw of MeCP2 protein and Histone H3 loading control from wild-type cells (WT), wild-type cells that went through the Nucleofection and cloning process (WTC), KO1 and KO2 cell lines. **(H)** Sequencing of sgRNA off-

target sites in KO1 and KO2 cell lines. Numbers next to each locus name indicate the off-target score as determined by crispr.mit.edu.

Insertion of Rett syndrome-causing point mutations into *MECP2*

We also explored the possibility of introducing specific point mutations into LUHMES cells, historically a more challenging procedure for somatic cells (Hendrie & Russell, 2005). The *MECP2* locus is an ideal candidate for use in optimising CRISPR knock-in (KI) conditions as there are a number of disease-causing point mutations throughout the locus (Lyst & Bird, 2015). Furthermore, the manipulation of this X-linked gene offers the opportunity to explore the ability of the CRISPR/Cas9 system to genetically manipulate genes on the inactive X chromosome.

In the previous experiment serial dilution was used to generate single cell colonies but we found that this method led to low efficiency of cloning and some colonies were derived from more than one genetically modified cell line, as several expressed alleles were detected in cDNA sequencing (Supplementary Figure 4C). In order to improve clonal selection we used FACS sorting to cleanly isolate single cells into a 96-well plate. LUHMES cells were amenable to this manipulation, with approximately 50-60% of wells repopulating to produce single-cell colonies.

First, the ability to knock-in the Rett syndrome-causing missense mutation of arginine at position 306 to cysteine (R306C) was tested (Figure 20A). The R306C mutation itself (CGC→TGC) creates a novel target sequence for the restriction enzyme HpyCH4V (Figure 20B). This allowed for easy screening of genomic DNA from single-cell clones using a restriction fragment length polymorphism assay (RFLP), with the positive clones from this assay being confirmed by sequencing. Initially, a plasmid targeting vector containing 2044bp of homology was used to deliver the R306C point mutation and a silent PAM-abolishing mutation to prevent re-cutting of a recombined allele. Out of 191 single cell clones that were screened using the HpyCH4V RFLP assay, 0 appeared to be positive (experiment 1 in Table 10). Next a 110bp single-stranded oligodeoxynucleotide (ssODN) was used in combination with a non-complementary sgRNA that cuts 6bp away from R306, and in this instance 1 cell line out of 69 was positive, giving a KI efficiency of 1.6% (Figure 20A+C+D, experiment 2 in Table 10). Again analysis of this cell line demonstrated no off-target cutting (Figure 20E). An alternative sgRNA:ssODN pair where sgRNA 2 cuts 31bp away from R306, but the sgRNA and ssODN were complementary to one another was also tested (Figure 21A). This combination produced two positive R306C cell lines (KI efficiency of 2.9%, experiment 3 in

Table 10), but both cell lines contained downstream deletions at the site of the double-strand break (DSB) (Figure 21B). These results suggest that a large distance between the point mutation and the DSB could be more susceptible to error-prone recombination and therefore subsequent INDEL formation.

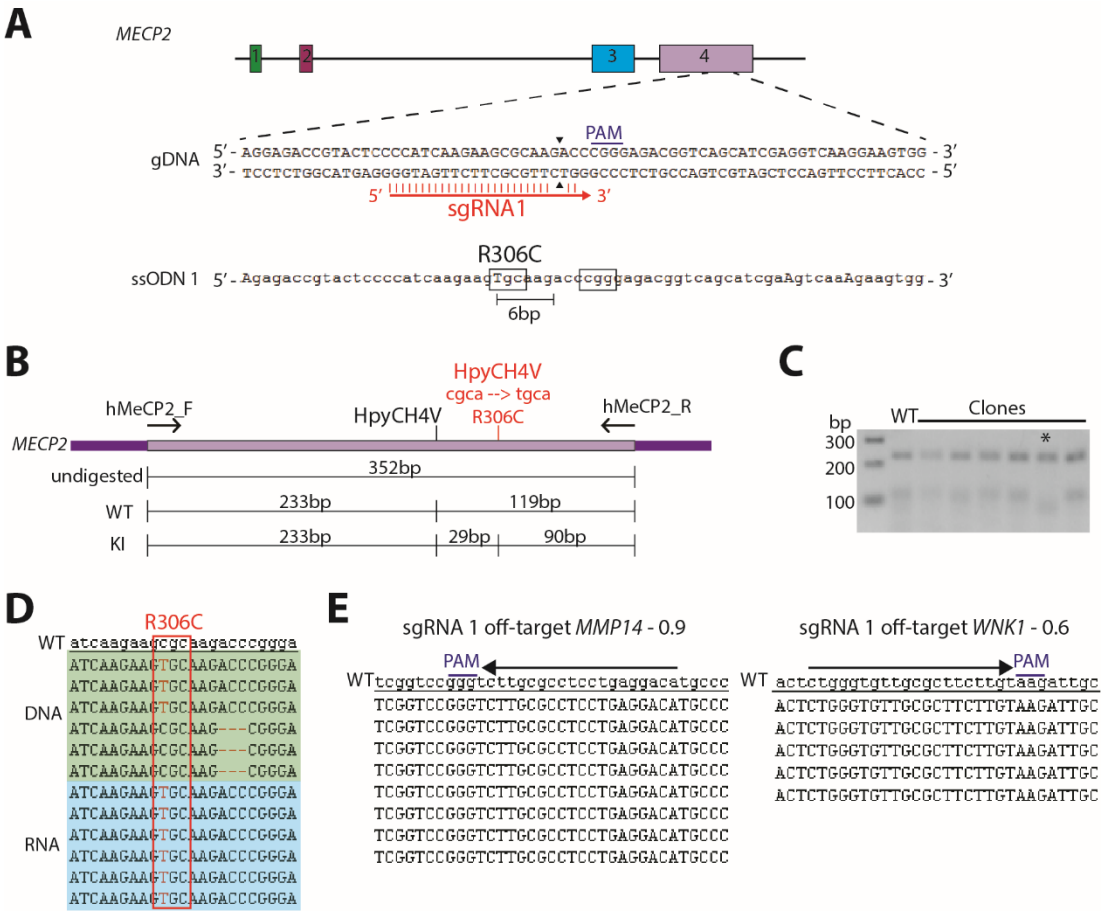


Figure 20. Generation of a human neuronal cell line containing a Rett syndrome-causing missense mutation in MECP2.

(A) Schematic representation of the *MECP2* locus with the sgRNA 1 target sequence labelled and the ssODN 1 donor molecule with point mutation alterations indicated in upper case. The site of double strand break is labelled with two arrowheads and the distance between the point mutation of interest and the double-strand break site is indicated. **(B)** Schematic representation of the RFLP screening assay used for identifying positive knock-in clones. Mutation of arginine at position 306 to cysteine results in the introduction of a novel target sequence for the restriction enzyme HpyCH4V. Primers used for PCR amplification are labelled. **(C)** HpyCH4V digests of the PCR product (Supp Fig 3A) to identify clones that have gained a novel HpyCH4V target sequence. A positive clone is identified with an asterisk. **(D)** Sequencing of genomic DNA and cDNA from the RFLP-positive cell line confirms the cell line to be *MECP2-R306C*. **(E)** Sequencing of the top two off-target sites for sgRNA 1 in the R306C cell line. Number next to the locus name is the off-target score as calculated by crispr.mit.edu.

In an attempt to increase the efficiency of KI the CRISPR plasmid that encodes puromycin resistance was exchanged for a CRISPR plasmid that encodes for green fluorescent protein (Ran *et al.*, 2013b) (Figure 21C). Thus, instead of subjecting the Nucleofected LUHMES

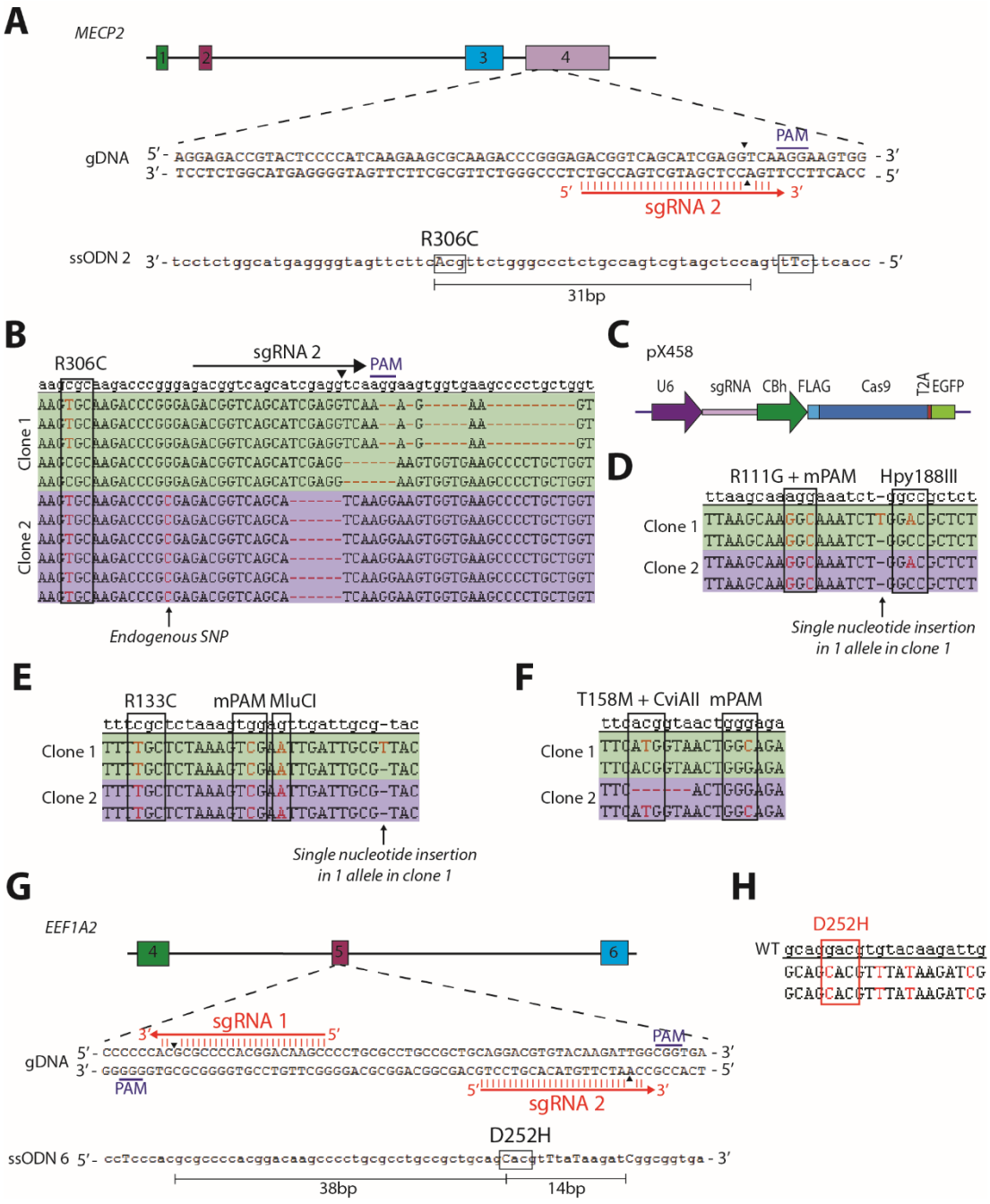


Figure 21. Generation of human neuronal cell lines containing missense mutations that cause neurological disorders.

(A) Schematic representation of the *MECP2* locus with the sgRNA 2 target sequence labelled. Site of double-strand break is indicated with arrowheads. ssODN molecule is shown with point mutation changes highlighted in upper case and the DSB to point mutation distance labelled. **(B)** Genomic DNA sequencing of two R306C positive clones. Both clones have the correctly inserted R306C point mutation as well as downstream deletions induced by sgRNA 2. The site of DSB is indicated by an arrowhead. **(C)** Schematic representation of a plasmid containing Cas9, sgRNA and EGFP expression constructs. **(D)** Genomic DNA sequencing of two R111G positive cell lines performed by Katie Paton. **(E)** Genomic DNA sequencing of two R133C positive cell lines performed by Katie Paton. **(F)** Genomic DNA sequencing of two T158M positive cell lines performed by Katie Paton. **(G)** Schematic representation of part of the *EEF1A2* locus, with sgRNA target sequences labelled and the ssODN donor molecule with point mutation alterations

indicated in upper case. Sites of single-strand nicks are indicated with an arrowhead. Distances between each nick and the point mutation of interest are labelled. Design of targeting experiment was performed by Dr Faith Davies. **(H)** Genomic DNA sequencing of the *EEF1A2-D252H* positive cell line performed by Dr Faith Davies.

cells to puromycin selection and FACS sorting, these two steps were combined into one by using the presence of GFP in cells two days after Nucleofection to identify positively transfected cells and to sort them into a 96-well plate. Using this new strategy three new Rett syndrome-causing point mutations, R111G, R133C and T158M, were targeted using ssODNs (Supplementary Figure 5). Two of these three point mutations do not introduce a novel restriction enzyme target sequence and so one was engineered into the ssODN for ease of screening. Each 100bp ssODN contained point mutations to introduce the following motifs: the mutation of interest, a silent PAM abolishing mutation, and a silent mutation to insert a novel restriction enzyme target sequence (Supplementary Figure 5). All three targeting experiments generated positive cell lines as judged by genomic DNA sequencing (Figure 21D,E,F) and as can be seen in Table 10, the efficiency of KI for all three point mutations is significantly increased relative to the initial efficiency of 1.6%, reaching a maximum of 26% for KI of R111G.

	Point mutation	Selection, plasmid amount	Donor molecule (ssODN is 10 µM)	sgRNA: ssODN	PM → DSB distance	# of US muts	# of KI clones	KI efficiency
1	R306C	Puro, 1.2 µg	1.9 µg 2 kb plasmid	-	31 bp	1	0/191	0%
2	R306C	Puro, 2 µg	4 µl 110 bp ssODN 1	Not comp	6 bp	1	1/69	1.6%
3	R306C	GFP, 2.5 µg	8 µl 110 bp ssODN 2	Comp	31 bp	1	2/69	2.9%
4	R111G	GFP, 2 µg	10 µl 100 bp ssODN 3	Comp	3 bp	0	7/27	26%
5	R133C	GFP, 2 µg	10 µl 100 bp ssODN 4	Comp	6 bp	2	2/54	3.7%
6	T158M	GFP, 2 µg	10 µl 100 bp ssODN 5	Comp	4 bp	1	1/18	5.5%
7	T158M	GFP, 2 µg	10 µl 100 bp ssODN 5	Comp	5 bp	1	1/13	7.7%

Table 10. Point mutation KI efficiencies in the *MECP2* locus.

PM = point mutation. DSB = double-strand break. # = number. US = upstream

As shown in Table 10, several factors could contribute to the large variability in targeting efficiency. Firstly, the distance of the sgRNA-induced DSB from the point mutation of interest varies, and secondly the number of mismatched residues in the ssODN that are upstream of the DSB also varies. This latter variable would be in line with evidence that mismatches in the non-sgRNA binding DNA strand upstream of the PAM are refractory to homology-directed repair (HDR) (Richardson *et al.*, 2016). Despite uncertainty regarding the exact constraints on efficient KI of point mutations using CRISPR technology, KI efficiencies in the somatic LUHMES neuronal progenitor cells are sufficient to allow the rapid generation of cell lines containing disease-causing point mutations with minimal clonal selection and screening.

Insertion of a neurodevelopmental disorder-causing point mutation into *EEF1A2*

To demonstrate the utility of LUHMES cells as a model system for other neurological disorders, and to confirm efficient KI at an autosomal locus, we targeted the D252H missense mutation in the *EEF1A2* gene that causes severe neurodevelopmental delay and intellectual disability (Nakajima *et al.*, 2015). The approach was to use two sgRNAs that cut 51bp apart, combined with Cas9 nickase protein (Ran *et al.*, 2013a) (Figure 21G) and a 200bp ssODN. Again the presence of GFP expression in LUHMES cells was used to identify positively transfected cells and to sort single cells into 96-well plates. In this experiment a KI efficiency of 14% was achieved and interestingly the D252H positive cell line has a KI on both alleles and is therefore a homozygous, true positive cell line (Figure 21H). The successful knock-in of a point mutation into an autosomal locus and subsequent generation of an *EEF1A2-D252H* cell line demonstrates the ease of genetic manipulation of LUHMES cells and highlights its utility for modelling a variety of human neurogenetic disorders.

Knock-in of a large tag into the endogenous *MECP2* locus

Finally, we assessed the ability of the CRISPR/Cas9 system to introduce a large tag into an endogenous locus in LUHMES cells. Again the *MECP2* locus was targeted and a sgRNA that spans the stop codon was used, resulting in its targeting sequence being abolished once a successful KI has occurred (Figure 22A). We chose to KI mCherry and use FACS analysis to provide an accurate estimate of KI efficiency, i.e. the percentage of mCherry positive cells in the whole population. Due to the large size of mCherry (711bp), a plasmid donor was used for targeting with 2.3kb and 1.2kb homology arms (Figure 22A). FACS analysis determined the percentage of mCherry positive cells in the entire population to be 0.015% (Figure 22B). Out of 29 single-cell clones assessed, 25 were *MECP2-mCherry* positive as judged by a PCR

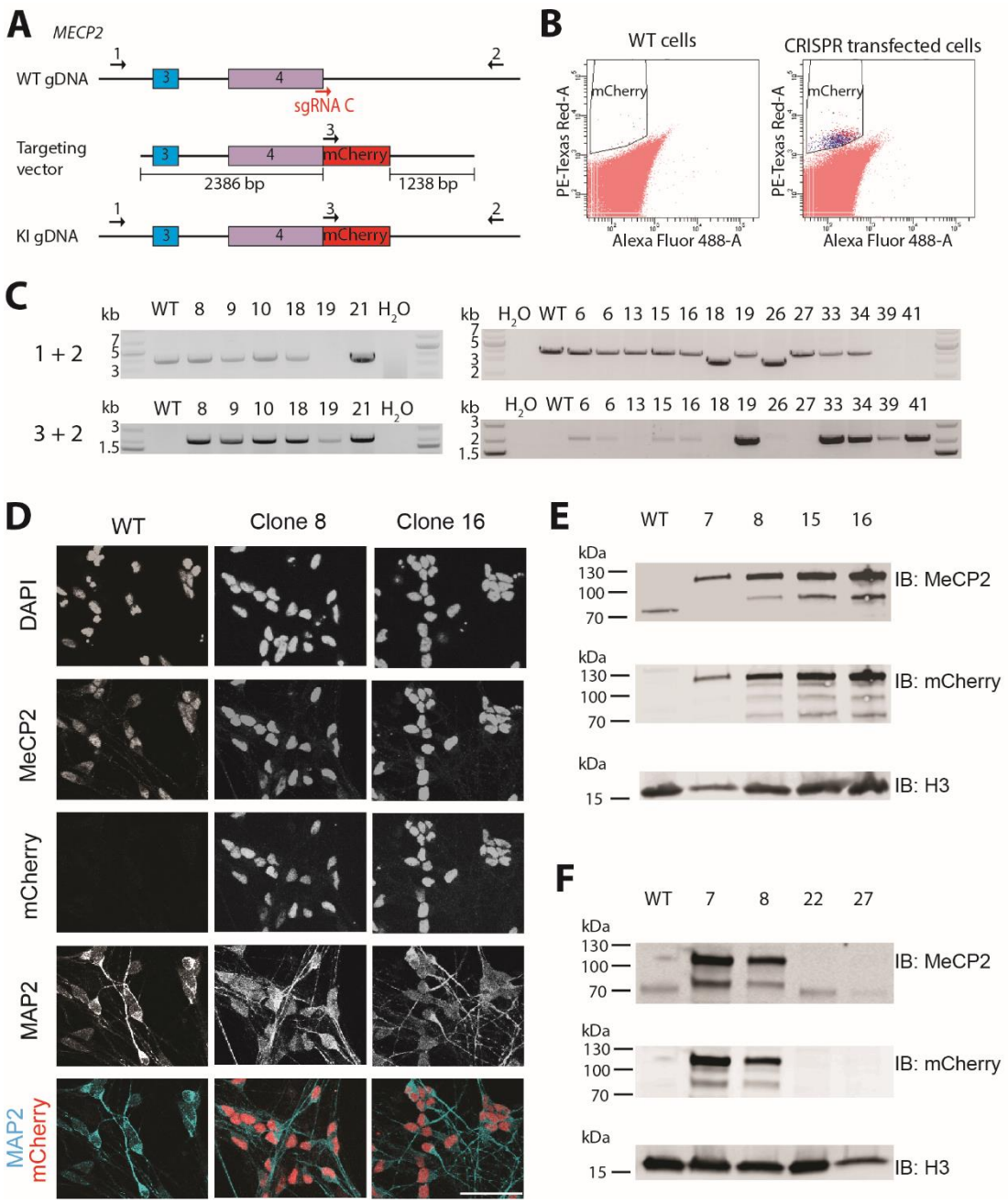


Figure 22. Endogenous knock-in of an mCherry tag into the MECP2 locus in LUHMES cells.

(A) Schematic representation of the *MECP2* locus with the sgRNA C target region labelled, the targeting vector, and the recombined genomic DNA allele. Positions of primers used for screening purposes in (C) are indicated. **(B)** Flow cytometry analysis of WT cells and cells that were transfected with Cas9, sgRNA and targeting plasmid. **(C)** PCR screening of genomic DNA from single-cell clones that were identified as being mCherry positive by flow cytometry. **(D)** Immunofluorescence imaging of WT cells and two single-cell clones using DAPI and antibodies probing for MeCP2 and MAP2. Images are slices through a z-stack. Scale bar represents 50 µm. **(E)** Immunoblot analysis of WT cells and four *MECP2-mCherry* positive clones. **(F)** Immunoblot analysis of WT cells, two *MECP2-mCherry* positive clones and two *MECP2-mCherry* negative clones. **(E+F)** Top panel probe: MeCP2. Middle panel probe: mCherry. Bottom panel probe: Histone H3 as a loading control.

assay that used a forward primer in mCherry itself and a reverse primer in the *MECP2* gene locus, outwith the targeting vector (Figure 22C). Positive cell lines were confirmed by immunofluorescence and Western blot analysis (Figure 22D+E +Supplementary Figure 6), and two negative cell lines (as determined by PCR analysis) were confirmed by Western blot (Figure 22F, clones 22 + 27). These experiments demonstrate successful CRISPR-mediated KI of a large tag into LUHMES cells, thus highlighting the variety of genetic alterations that are feasible in this cell line.

4.3.5 Discussion

The LUHMES cell line is an immortalised neuronal cell line derived from an 8-week old female foetus that is highly proliferative in a stem-cell like, yet neuronal-committed state, and can differentiate into mature dopaminergic neurons via addition of tetracycline to the cell culture medium (Scholz *et al.*, 2011). A key advantage of LUHMES cells compared to other neuronal differentiation systems is the near 100% homogeneity of differentiation into a population of mature, post-mitotic neurons, without the presence of astrocytes or other non-neuronal cell types. This homogeneity is extremely beneficial for "bulk population" experiments such as RNA-sequencing, Western blot analysis and Hi-C studies where mixed cell populations could result in skewed data and difficult-to-interpret results. Here we describe methods for the successful genetic manipulation of LUHMES cells in order to create targeted protein knock-out, disease-causing point mutation knock-in, and large tag knock-in cell lines.

The combination of targeted mutagenesis with rapid generation of mutant neuronal cells provides a potentially valuable tool for neuroscience. These manipulated cell lines may complement *in vivo* datasets as disease phenotypes obtained using mouse models can be coupled with electrophysiological and biochemical data from human neurons in order to bridge the gaps between disease causing mutations, neuronal malfunction and whole organism pathophysiology. The ease of high-throughput differentiation of this cell line in 96-well plates opens the door towards drug screening programs like those already being pursued using hiPSCs (Cao *et al.*, 2016; Lee *et al.*, 2012). As a potential alternative, LUHMES cells simplify the differentiation procedure, speed up the differentiation time course, and ensure that a homogeneous population of mature neurons will be screened. Furthermore, although downstream applications of genetically modified LUHMES cells might be limited by their dopaminergic lineage, removal of supplements from the differentiation medium results in the

production of tyrosine hydroxylase-negative cells, that differentiate into morphologically and immunocytologically mature neurons for study (Scholz *et al.*, 2011).

Bi-allelic X chromosome targeting

We observed that both alleles undergo HDR at a rather high frequency, regardless of whether the 2nd allele is on an autosome or the inactive X chromosome. For the *MECP2-R111G* targeting experiment, out of the seven cell lines that contained a KI of R111G on the active X allele, six also contained a KI on the inactive X allele. Likewise, for the *MECP2-R133C* targeting experiment, the two R133C positive cell lines had undergone HDR repair on both alleles. It is however important to note that not all HDR events result in a clean integration. Partial recombination within the short distance of a 100bp ssODN (Figure 23A+B) and multiple integrations of the ssODN in tandem at a locus have been observed (Figure 23E), as well as recombined alleles containing INDELs at the site of sgRNA DSB (Figure 21B).

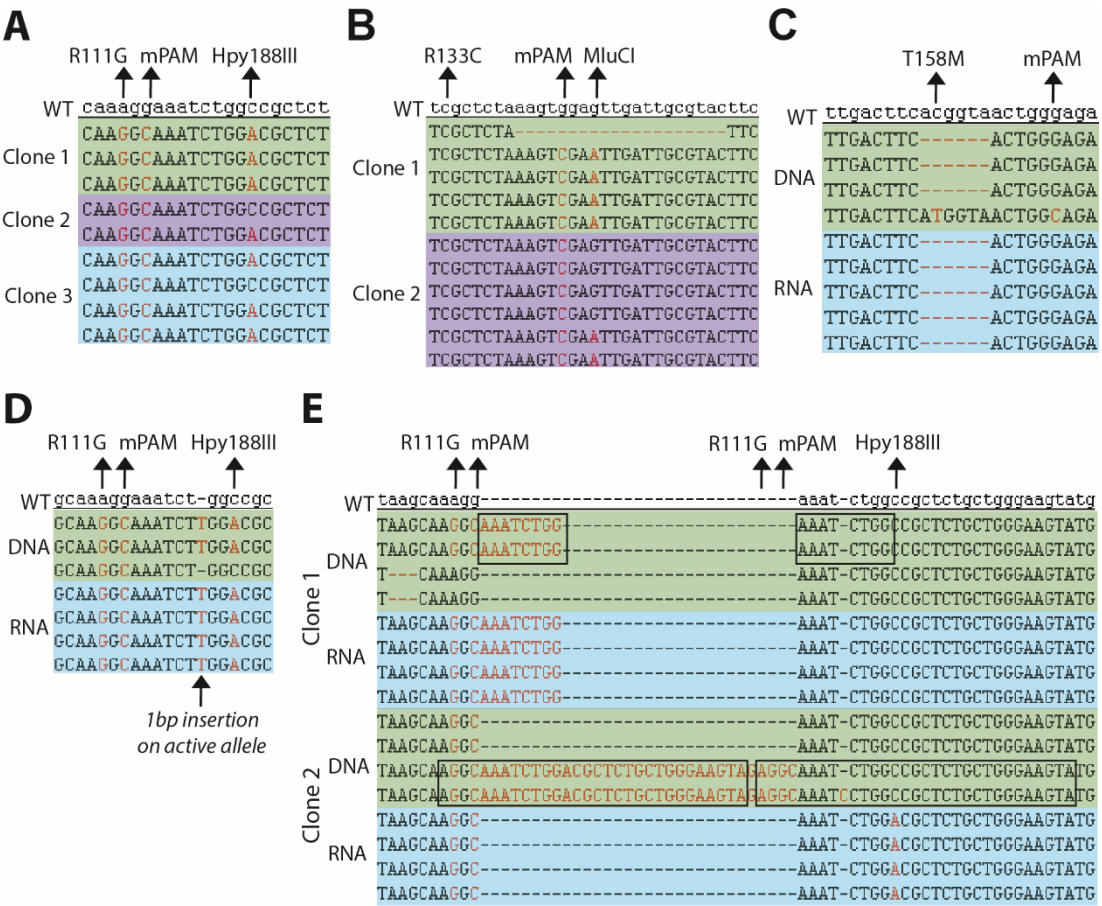


Figure 23. Analysis of range of mutations induced by CRISPR/Cas9 in LUHMES cells.

(A+B) Partial recombination can occur with the ssODN molecule. **(A)** Three single cell clones from the R111G targeting experiment, two of which contain a partially recombined allele. Sequencing performed by Katie Paton. **(B)** Two single cell clones from the R133C targeting experiment; clone 1 has an allele with 2/3 of the KI residues, clone 2 has an allele with only 1/3 of the KI residues. Sequencing performed b Katie Paton. **(C)** Genomic DNA sequencing of a single clone reveals one allele to contain two point mutations and the other allele to contain a 6bp deletion, yet cDNA sequencing identifies the HDR to have occurred with the allele on the inactive X chromosome. DNA sequencing performed by Katie Paton. **(D)** cDNA sequencing identifies the allele on the active X chromosome having undergone recombination, with all three point mutations integrated plus a single base pair insertion. DNA sequencing performed by Katie Paton. **(E)** Sequencing of single-cell clones reveals duplication events occurring as a result of multiple recombination events with the ssODN. DNA sequencing performed by Katie Paton. **(A,B,C,D,E)** mPAM - silent point mutation that abolishes the PAM site for each sgRNA. Rett-syndrome causing point mutations of interest and silent restriction enzyme target sequence point mutations are highlighted with arrows.

Surprisingly we found in more than one case that the *MEPC2* allele on the inactive X chromosome underwent HDR, while the allele on the active X chromosome acquired an INDEL. It is expected that the active allele would be more open and accessible to recombination compared to the inactive X chromosome, however at least two cell lines were observed that have an active allele INDEL and an inactive allele KI (Figure 23C+D). These data reflect the large variety of genomic alterations that can be induced by CRISPR/Cas9 and demonstrate the somewhat unpredictable nature of HDR-mediated gene targeting. Our experiments stress the need for a sgRNA-induced DSB as close as possible to the desired genetic alteration and highlight the importance of donor molecule design; in particular use of a sgRNA-complementary ssODN that has minimal mismatches upstream of the PAM seems to be most efficient.

Further optimisation of KI of large tags

In the mCherry KI experiment the plentiCRISPR plasmid (Sanjana *et al.*, 2014) was used to deliver Cas9 and sgRNA, and positively transfected cells were selected using the co-encoded puromycin resistance gene (Figure 19B), while the mCherry targeting vector was delivered as a separate plasmid. It is possible that a proportion of transfected cells did not take up both plasmids (Assur *et al.*, 2012) and this could explain the low targeting efficiency of 0.015% observed in this experiment. As such, a double antibiotic selection method could increase HDR efficiencies when plasmid donors are necessary, for example by including an expression cassette for the bacterial blasticidin resistance gene (bsr) in the targeting plasmid and selecting with both puromycin and blasticidin.

Design of targeting vectors with alternative homology arm lengths could also improve the efficiency of KI. Indeed, others have reported comparable efficiencies with homology arms of 175bp compared to 700-900bp in human cell lines (Natsume *et al.*, 2016) Alternatively, the use of drugs to inhibit the NHEJ pathway could also boost HDR in LUHMES cells (Maruyama *et al.*, 2015; van Overbeek *et al.*, 2016). Even without these enhancements, the power of FACS sorting allows efficient selection for the small number of positive cells within a large population and, as demonstrated here, this results in a stream-lined and efficient protocol for CRISPR-mediated tag KI in LUHMES cells.

Concluding remarks

In conclusion, we have demonstrated efficient genetic manipulation of the LUHMES female human neuronal cell line to create a number of lines harbouring neurological disease-causing point mutations (
Figure 24). The future phenotypic assessment of these cell lines will provide significant insight into the molecular mechanisms of these diseases. Using the methods described here LUHMES cells have the potential to be a valuable tool for exploration of the underlying biology of neurogenetic disorders and may pave the way for drug development and therapeutic strategies in the future.

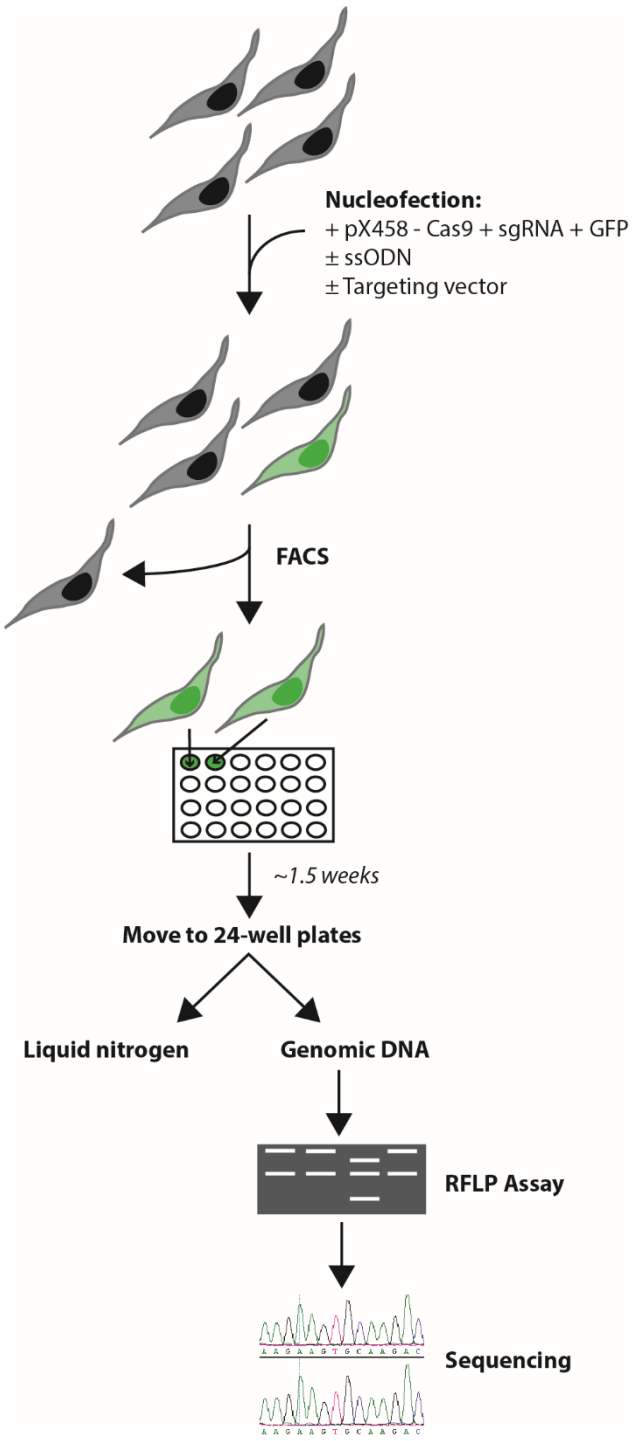


Figure 24. Simple targeting pipeline for generation of genetically modified LUHMES cell lines using CRISPR technology.

Cells are Nucleofected with the plasmids and ssODNs necessary for the specific targeting experiment and taken for FACS sorting two days later. After approximately 10 days single cell colonies will have expanded enough to require dissociation and transfer to 24-well plates. From 24-well plates cells can be frozen down for storage in liquid nitrogen and some cells harvested for genomic DNA. The first step of the screening strategy is to perform an RFLP assay to identify a subset of clones that will be taken for genomic DNA sequencing which will identify cell lines that are positive for a clean KI. For more detailed information see methods and materials.

4.4 MeCP2 protein level in CRISPR-mutated cell lines

Supplementary Figure 4B in Chapter 9.5 demonstrates that the insertion of INDELs into exon 3 of the *MECP2* gene locus can result in the KO of MeCP2 protein from LUHMES cells. The other targeting experiments from Chapter 4.3 generated not only positive, point mutation-containing cell lines such as R306C and R111G, but also a variety of other cell lines containing INDELs of different sizes. There are two questions here: does the insertion of R306C or R111G point mutations have any effect on the stability of the MeCP2 protein, and to what extent do other INDELs in exons 3 and 4 result in KO or truncation of MeCP2 protein? The latter question is useful because understanding the effects of random INDELs on MeCP2 protein stability will help to guide future experiments which aim to use CRISPR to fix particular RTT-causing mutations in MeCP2.

Figure 25A shows cDNA sequencing from the R306C cell line (2_9) and four other cell lines that were generated from this targeting experiment (Table 11, experiment number 2). This experiment used a sgRNA that targeted the NID of MeCP2, which is encoded by exon 4. The effects of the 5bp insertion, 35bp deletion and 27bp insertion on the nucleotide sequence of *MECP2* are schematically illustrated in Figure 25B, with the two out-of-frame INDELs resulting in an early stop codon after 11 or 15 nonsense residues, which would be predicted to produce a truncated protein. To assess what effect these INDELs have on the level of MeCP2 protein I performed a Western blot (Figure 25C+D). Panel C probes for the N-terminus of MeCP2 and thus picks up C-terminally truncated proteins, which are not detected by the C-terminal antibody in panel D. As predicted, the 4_5 and 4_6 cell lines have truncated forms of MeCP2. These proteins run at the predicted ~38kDa mark for their 317 or 321 amino acid size (Figure 25C). The properties that cause full-length WT MeCP2 protein to run at ~80kDa instead of its expected 53kDa size is not present in these C-terminally truncated proteins. Both cell lines have a reduced level of MeCP2 compared to WT neurons, yet it is surprising that 4_6 is more unstable than 4_5 when there is only a four amino acid difference in the size of these two proteins. Could the gain of four extra amino acids be enough to stabilise a truncated MeCP2? Perhaps more likely, the deletion of 35bp in the RNA species results in decreased levels of mRNA either through reduced transcription and processing of transcripts or due to inherent instability of the RNA species itself. The in-frame insertion of 27bp in cell line 4_18 is predicted to produce a protein that has WT sequence until residue 304, with a large nonsense tail afterwards. This causes a complete KO of all MeCP2 protein in this cell line (Figure 25C+D). Finally, the 4_20 cell line has extremely low levels of a truncated protein (runs just higher than MeCP2 in 4_5 and 4_6,

Figure 25C), yet cDNA sequencing from this cell line was WT (Figure 25A). In order to establish the precise mutation in this cell line that results in such an unstable, truncated protein, further sequencing of genomic DNA and cDNA should be performed.

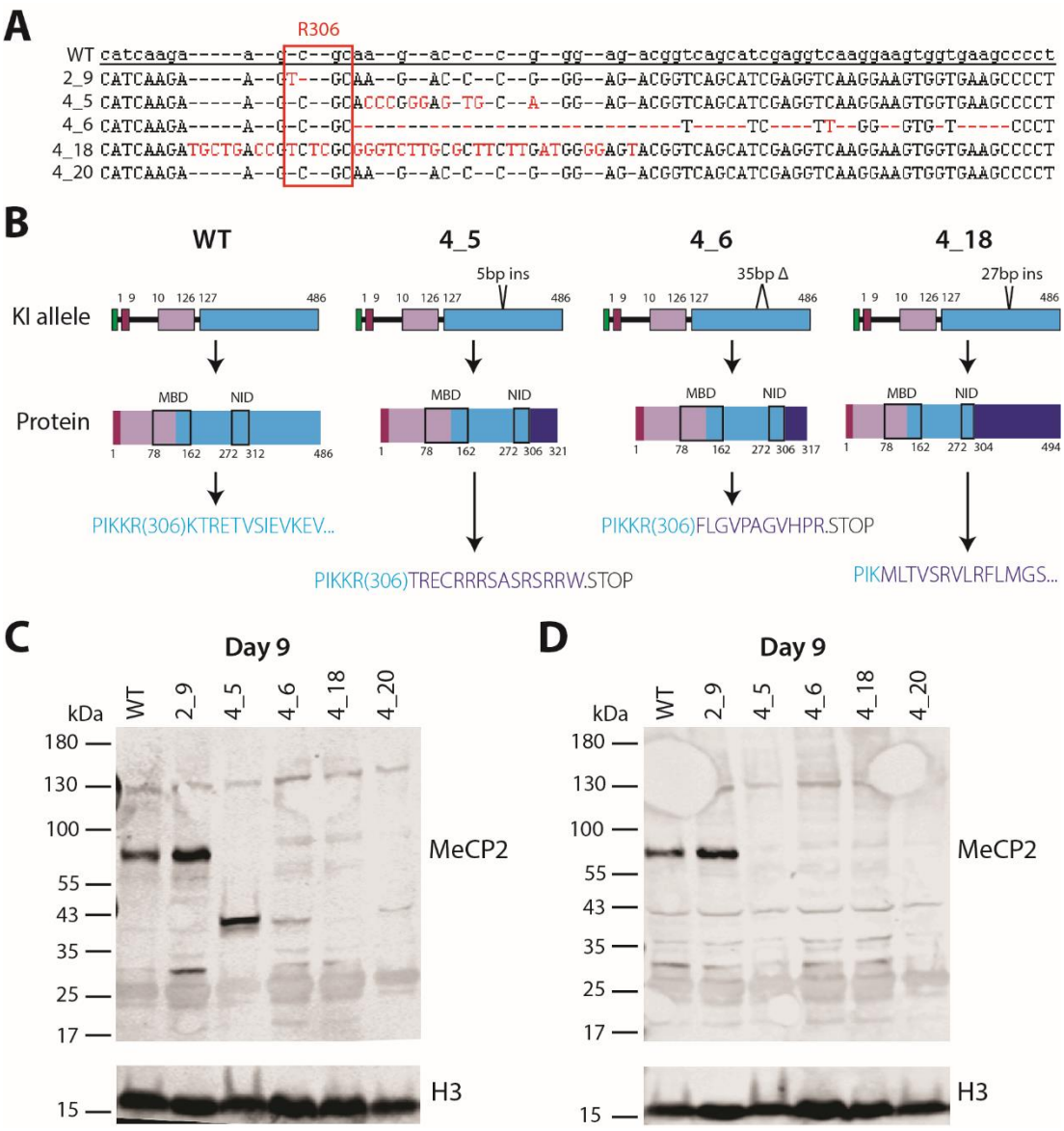


Figure 25. Sequence and Western blot analysis of CRISPR-modified cell lines from the R306C targeting experiment.

(A) Sequencing of cDNA from 2_9 (R306C-positive) cell line, and 4 other lines from the same targeting experiment. **(B)** Schematic illustrations of cDNA and protein sequences in non-R306C cell lines. Light blue box and writing indicates WT amino acid sequence encoded from exon 4. Dark blue box and writing indicates CRISPR-induced mutations. MBD amino acids as defined in Nan et al., 2003 and NID amino acids as defined in Lyst et al., 2013. **(C+D)** Western blots of WT cells and cell lines derived from the R306C targeting experiment. 2_9 = R306C KI. Anti-MeCP2 antibody is M7443 (C) from Sigma which detects the N-terminal region of the protein, coded from the beginning of exon 3 or M6818 (D) from Sigma which detects the C-terminal region of the protein, coded from the end of exon 4. Western blots are representative images from 5 replicates.

To explore the effect of MBD-containing INDELs on MeCP2 protein level, clonal cell lines from targeting experiments 3 and 6 (

Experiment	Clone	DNA sequence		RNA sequence	Protein
		allele 1	allele 2		
	H4 (KO2)	14bp del	WT	14bp del	KO
	D10 (KO1)	6bp del	6bp del	Exon 3 removed	KO
1. KO	H5*	1bp ins	11bp ins or WT	1bp ins	KO
sgRNA 16 or 17	C12	1bp ins or WT	7bp del or 6bp del	1bp ins + 7bp del	Practically KO
(puro selection)	F9	2bp ins	27bp del	2bp ins + 27bp del	Reduced
	C11	6bp del	7bp del	WT or 7bp del	Reduced
	A11			WT	Equal
	2_9	R306C	3bp del	R306C	Equal
2. R306C	4_5*	5bp ins	5bp ins	5bp ins	Truncated, equal
sgRNA 15	4_6	WT	35bp del	35bp del	Truncated, reduced
ssODN 1	4_18*	27bp ins	27bp ins	27bp ins	KO
(puro selection)	4_20*			WT	Truncated, v. low
	2_7	3 bp del	8 bp ins	8 bp ins	KO
	2_13	R111G; no RE	R111G; RE	R111G ± RE	Reduced
3. R111G	2_17	R111G; RE	R111G; RE	R111G ± RE	Reduced
sgRNA K2	2_20	R111G; RE; 1 bp ins	R111G; no RE	R111G + 1bp ins	KO
ssODN 4	2_21	R111G; RE; 2 bp sub	R111G; no RE	R111G ± RE	Reduced
	2_22	31bp ins	R111G; no RE	R111G	Reduced
	2_30	R111G; RE	R111G; no RE	R111G ± RE	Reduced
	2_31	R111G; RE	R111G; no RE; 1bp ins	R111G ± RE	Reduced
	4_23	7bp del	12 bp del		
4. R133C	4_28				KO
sgRNA K4	4_37*	19bp del	19bp del		
ssODN 5	4_40	No R133C; RE; mPAM	30bp del		
	4_47	R133C; RE; mPAM	R133C; RE; mPAM	DIED	DIED
5. T158M					
sgRNA K5	5_20	T158M; mPAM	WT	DIED	DIED
ssODN 6					
6. T158M					
sgRNA K6	6_48	6bp del	T158M; mPAM	6bp del	Reduced
ssODN 6					
7. R306C					
sgRNA 5	4		R306C; mPAM; 6bp del	R306C; mPAM; 6bp d	
ssODN 3					
	1G10	R133C; RE; mPAM	R133C; RE; mPAM; 1bp ins	R133C; RE; mPAM	
8. R133C	2E8			24bp del	
sgRNA K4	2D8			GT to AGG	
ssODN 5	2D5			2bp del	
	2E6			7bp ins	
9. R306C	2D4	4bp del	R306C; 18bp del	R306C; 18bp del	
sgRNA 5	2H9			1bp ins	
ssODN 2	1E3			1bp del	

Table 11) were assessed by Western blot analysis. Each experiment used a sgRNA whose target sequence was within the MBD, but experiments 3 and 4 targeted exon 3 of *MECP2* and experiment 6 targeted exon 4 of *MECP2*. Cell lines 2_7 and 2_20, which were both generated from experiment 3, contain an 8bp and a 1bp insertion respectively (Figure 26A, Table 11). Both of these out-of-frame insertions are predicted to produce a truncated form of MeCP2 protein (

Figure 26A). On the other hand, cell line 6_48, which was produced from experiment 6, contains an in-frame, 6bp deletion which deletes Threonine 158 and Valine 159 (Figure 26B,

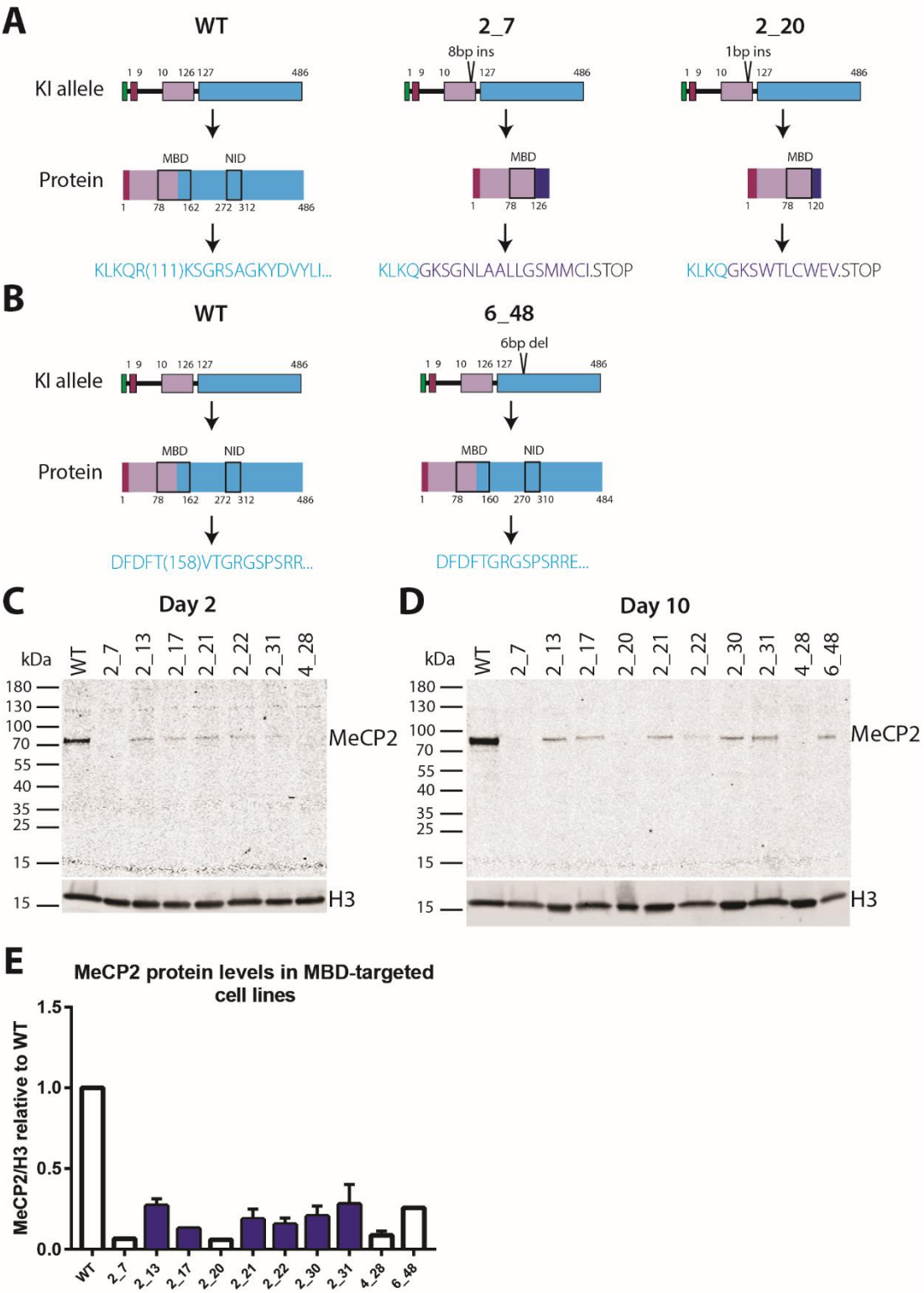
Experiment	Clone	DNA sequence		RNA sequence	Protein
		allele 1	allele 2		
	H4 (KO2)	14bp del	WT	14bp del	KO
	D10 (KO1)	6bp del	6bp del	Exon 3 removed	KO
1. KO	H5*	1bp ins	11bp ins or WT	1bp ins	KO
sgRNA 16 or 17	C12	1bp ins or WT	7bp del or 6bp del	1bp ins + 7bp del	Practically KO
(puro selection)	F9	2bp ins	27bp del	2bp ins + 27bp del	Reduced
	C11	6bp del	7bp del	WT or 7bp del	Reduced
	A11			WT	Equal
	2_9	R306C	3bp del	R306C	Equal
2. R306C	4_5*	5bp ins	5bp ins	5bp ins	Truncated, equal
sgRNA 15	4_6	WT	35bp del	35bp del	Truncated, reduced
ssODN 1	4_18*	27bp ins	27bp ins	27bp ins	KO
(puro selection)	4_20*			WT	Truncated, v. low
	2_7	3 bp del	8 bp ins	8 bp ins	KO
	2_13	R111G; no RE	R111G; RE	R111G ± RE	Reduced
3. R111G	2_17	R111G; RE	R111G; RE	R111G ± RE	Reduced
sgRNA K2	2_20	R111G; RE; 1 bp ins	R111G; no RE	R111G + 1bp ins	KO
ssODN 4	2_21	R111G; RE; 2 bp sub	R111G; no RE	R111G ± RE	Reduced
	2_22	31bp ins	R111G; no RE	R111G	Reduced
	2_30	R111G; RE	R111G; no RE	R111G ± RE	Reduced
	2_31	R111G; RE	R111G; no RE; 1bp ins	R111G ± RE	Reduced
	4_23	7bp del	12 bp del		
4. R133C	4_28				KO
sgRNA K4	4_37*	19bp del	19bp del		
ssODN 5	4_40	No R133C; RE; mPAM	30bp del		
	4_47	R133C; RE; mPAM	R133C; RE; mPAM	DIED	DIED
5. T158M					
sgRNA K5	5_20	T158M; mPAM	WT	DIED	DIED
ssODN 6					
6. T158M					
sgRNA K6	6_48	6bp del	T158M; mPAM	6bp del	Reduced
ssODN 6					
7. R306C					
sgRNA 5	4		R306C; mPAM; 6bp del	R306C; mPAM; 6bp d	
ssODN 3					
	1G10	R133C; RE; mPAM	R133C; RE; mPAM; 1bp ins	R133C; RE; mPAM	
8. R133C	2E8			24bp del	
sgRNA K4	2D8			GT to AGG	
ssODN 5	2D5			2bp del	
	2E6			7bp ins	
9. R306C	2D4	4bp del	R306C; 18bp del	R306C; 18bp del	
sgRNA 5	2H9			1bp ins	
ssODN 2	1E3			1bp del	

Table 11**).** This should produce a protein lacking only two amino acids and therefore is 484 amino acids long (Figure 26B). Threonine 158 is a critical residue for structural stability in

the MBD (Ho *et al.*, 2008), thus it would be expected that the loss of this residue has a dramatic effect on the stability of MeCP2. In line with this, a T158M mutant mouse line also exhibits decreased protein levels (Brown *et al.*, 2016). Western blot analysis demonstrated that the two out of frame INDELs in exon 3 result in a complete KO of MeCP2 protein (2_7 + 2_20, Figure 26C+D+E), while cell line 6_48, as expected, has protein that runs at the same size as WT protein, but has reduced stability (Figure 26D+E).

Figure 26. Western blot analysis of CRISPR-modified cell lines from the R111G, R133C and T158M targeting experiments.

(A+B) Schematic illustrations of cDNA and protein sequences in the R111G (A) and T158M (B) targeting experiments. Light blue box and writing indicates WT amino acid sequence encoded from exon 4. Dark blue box and writing indicates CRISPR-induced mutations. MBD amino acids as defined in Nan et al, 2003 and NID amino acids as defined in Lyst et al 2013. **(C+D)** Western blots of WT and MBD-targeted cell lines at day 2 (C) or day 10 (D) of differentiation. **(E)** Quantification of MeCP2/H3 ratio in R111G KI and MeCP2 KO 10-day old LUHMES-derived neurons. Blue bars are R111G KI cell lines. N=1-2. Antibody in C+D+E is M7443, Sigma.


A
WT
2_7
2_20
8bp ins
1bp ins
KI allele
Protein
MBD
NID
KLKQR(111)KSGRSAGKYDVYLI...
KLKQGKSGNLAALLGSMMCI.STOP
KLKQGKSWTLCWEV.STOP
B
WT
6_48
6bp del
KI allele
Protein
DFDFT(158)VTGRGSPSRR...
DFDFTGRGSPSRRE...
C
Day 2
kDa
WT
2_7
2_13
2_17
2_21
2_22
2_31
4_28
MeCP2
H3
D
Day 10
2_20
2_30
6_48
E
MeCP2 protein levels in MBD-targeted cell lines
MeCP2/H3 relative to WT

While I am cautious to draw too many conclusions from observational studies of a small number of clonal cell lines, the data suggests that an out-of-frame mutation in the MBD results in an early stop codon and a complete KO of protein (2_7, 2_20, KO2, H5) while an in-frame mutation is tolerated but results in decreased levels of protein, possibility due to instability of the protein itself (6_48). Whereas the opposite is true for mutations in the NID, where an out-of-frame mutation produces an early stop codon resulting in reduced levels of a truncated protein (4_5 and 4_6), while an in-frame mutation is not tolerated and results in a complete KO of protein (4_18). This conclusion is somewhat hypothetical and analysis of other LUHMES clonal cell lines described in

Experiment	Clone	DNA sequence		RNA sequence	Protein
		allele 1	allele 2		
	H4 (KO2)	14bp del	WT	14bp del	KO
	D10 (KO1)	6bp del	6bp del	Exon 3 removed	KO
1. KO	H5*	1bp ins	11bp ins or WT	1bp ins	KO
sgRNA 16 or 17	C12	1bp ins or WT	7bp del or 6bp del	1bp ins + 7bp del	Practically KO
(puro selection)	F9	2bp ins	27bp del	2bp ins + 27bp del	Reduced
	C11	6bp del	7bp del	WT or 7bp del	Reduced
	A11			WT	Equal
	2_9	R306C	3bp del	R306C	Equal
2. R306C	4_5*	5bp ins	5bp ins	5bp ins	Truncated, equal
sgRNA 15	4_6	WT	35bp del	35bp del	Truncated, reduced
ssODN 1	4_18*	27bp ins	27bp ins	27bp ins	KO
(puro selection)	4_20*			WT	Truncated, v. low
	2_7	3 bp del	8 bp ins	8 bp ins	KO
	2_13	R111G; no RE	R111G; RE	R111G ± RE	Reduced
3. R111G	2_17	R111G; RE	R111G; RE	R111G ± RE	Reduced
sgRNA K2	2_20	R111G; RE; 1 bp ins	R111G; no RE	R111G + 1bp ins	KO
ssODN 4	2_21	R111G; RE; 2 bp sub	R111G; no RE	R111G ± RE	Reduced
	2_22	31bp ins	R111G; no RE	R111G	Reduced
	2_30	R111G; RE	R111G; no RE	R111G ± RE	Reduced
	2_31	R111G; RE	R111G; no RE; 1bp ins	R111G ± RE	Reduced
	4_23	7bp del	12 bp del		
4. R133C	4_28				KO
sgRNA K4	4_37*	19bp del	19bp del		
ssODN 5	4_40	No R133C; RE; mPAM	30bp del		
	4_47	R133C; RE; mPAM	R133C; RE; mPAM	DIED	DIED
5. T158M					
sgRNA K5	5_20	T158M; mPAM	WT	DIED	DIED
ssODN 6					
6. T158M					
sgRNA K6	6_48	6bp del	T158M; mPAM	6bp del	Reduced
ssODN 6					
7. R306C					
sgRNA 5	4		R306C; mPAM; 6bp del	R306C; mPAM; 6bp d	
ssODN 3					
	1G10	R133C; RE; mPAM	R133C; RE; mPAM; 1bp ins	R133C; RE; mPAM	
8. R133C	2E8			24bp del	
sgRNA K4	2D8			GT to AGG	
ssODN 5	2D5			2bp del	
	2E6			7bp ins	
9. R306C	2D4	4bp del	R306C; 18bp del	R306C; 18bp del	
sgRNA 5	2H9			1bp ins	
ssODN 2	1E3			1bp del	

Table 11 would be necessary to assess this. Larger-scale and more rigorous studies into the exact nature of INDEL consequences on MeCP2 protein stability would also be beneficial,

especially if approaches using CRISPR technologies are to be explored as methods for correction of mutant *MECP2* alleles in RTT.

Turning now to the RTT missense mutation cell lines, the R306C point mutation in the NID of MeCP2 does not result in a destabilised protein (Figure 25C+D). This is as predicted based on a previous study which showed that an EGFP-tagged R306C KI mouse line has equal protein levels when compared to an EGFP-tagged WT KI mouse line (Brown *et al.*, 2016). On the other hand, the R111G point mutation results in reduced levels of MeCP2 protein that can be observed after only 2 days of differentiation (

Figure 26C+D+E). While the effect of R111G on methylated DNA binding is well established (Free *et al.*, 2001; Nan *et al.*, 2007), its effect on protein stability has not been thoroughly assessed. The only *in vivo* model of this point mutation consists of a transgenic mouse line in a null background where R111G-MeCP2 is tagged with EGFP. Unfortunately, the transgenic nature of the *R111G-MECP2* allele means that allele copy number from multiple lentiviral integrations will impact the protein level in this mouse line (Heckman *et al.*, 2014). *In vitro* assessments have found the R111G point mutation to affect the stability of the MBD (residues 77-167), although the results are somewhat contradictory. Two-dimensional nuclear magnetic resonance (NMR) spectroscopy of the MBD of MeCP2 found local structural perturbations in the region of 110-120 upon mutation of arginine 111 to glycine, but concluded the MBD to have no gross structural changes that would impact stability (Free *et al.*, 2001). On the other hand, urea-induced unfolding circular dichroism (CD) assays found the R111G MBD to have enhanced stability compared to WT MBD (Yang *et al.*, 2016). Furthermore, in a single paper two alternative conclusions were made about MBD stability changes induced by R111G. *In silico* modelling approaches

hypothesised that R111G would significantly destabilise the MBD, while urea-induced denaturation CD experiments found a stabilising effect (Kucukkal *et al.*, 2015).

To conclude, the R306C point mutation in LUHMES-derived neurons, as predicted, does not alter the protein level of MeCP2. The assessment of the R111G point mutation here is the first experiment *in vivo* to look at the effect of this mutation on MeCP2 protein levels and found that R111G-MeCP2 cell lines have approximately 20-30% protein in human neurons. The effect of other variable insertion and deletion mutations on MeCP2 protein level seems to depend on the nature of the INDEL itself (in-frame vs out-of-frame) and on the location of the mutation within the MeCP2 molecule.

4.5 Clonal cell line summary

Six different genetic modifications of the *MECP2* locus have been generated in LUHMES cell lines (KO of protein, KI of four point mutations and KI of a mCherry tag,

Figure 27) over nine different targeting experiments.

Experiment	Clone	DNA sequence		RNA sequence	Protein
		allele 1	allele 2		
	H4 (KO2)	14bp del	WT	14bp del	KO
	D10 (KO1)	6bp del	6bp del	Exon 3 removed	KO
1. KO	H5*	1bp ins	11bp ins or WT	1bp ins	KO
sgRNA 16 or 17	C12	1bp ins or WT	7bp del or 6bp del	1bp ins + 7bp del	Practically KO
(puro selection)	F9	2bp ins	27bp del	2bp ins + 27bp del	Reduced
	C11	6bp del	7bp del	WT or 7bp del	Reduced
	A11			WT	Equal
	2_9	R306C	3bp del	R306C	Equal
2. R306C	4_5*	5bp ins	5bp ins	5bp ins	Truncated, equal
sgRNA 15	4_6	WT	35bp del	35bp del	Truncated, reduced
ssODN 1	4_18*	27bp ins	27bp ins	27bp ins	KO
(puro selection)	4_20*			WT	Truncated, v. low
	2_7	3 bp del	8 bp ins	8 bp ins	KO
	2_13	R111G; no RE	R111G; RE	R111G ± RE	Reduced
3. R111G	2_17	R111G; RE	R111G; RE	R111G ± RE	Reduced
sgRNA K2	2_20	R111G; RE; 1 bp ins	R111G; no RE	R111G + 1bp ins	KO
ssODN 4	2_21	R111G; RE; 2 bp sub	R111G; no RE	R111G ± RE	Reduced
	2_22	31bp ins	R111G; no RE	R111G	Reduced
	2_30	R111G; RE	R111G; no RE	R111G ± RE	Reduced
	2_31	R111G; RE	R111G; no RE; 1bp ins	R111G ± RE	Reduced
	4_23	7bp del	12 bp del		
4. R133C	4_28				KO
sgRNA K4	4_37*	19bp del	19bp del		
ssODN 5	4_40	No R133C; RE; mPAM	30bp del		
	4_47	R133C; RE; mPAM	R133C; RE; mPAM	DIED	DIED
5. T158M					
sgRNA K5	5_20	T158M; mPAM	WT	DIED	DIED
ssODN 6					
6. T158M					
sgRNA K6	6_48	6bp del	T158M; mPAM	6bp del	Reduced
ssODN 6					
7. R306C					
sgRNA 5	4		R306C; mPAM; 6bp del	R306C; mPAM; 6bp d	
ssODN 3					
	1G10	R133C; RE; mPAM	R133C; RE; mPAM; 1bp ins	R133C; RE; mPAM	
8. R133C	2E8			24bp del	
sgRNA K4	2D8			GT to AGG	
ssODN 5	2D5			2bp del	
	2E6			7bp ins	
9. R306C	2D4	4bp del	R306C; 18bp del	R306C; 18bp del	
sgRNA 5	2H9			1bp ins	
ssODN 2	1E3			1bp del	

Table 11 outlines the genetic structure of all clones generated from these targeting experiments, in order to provide an easy-to-reference table.

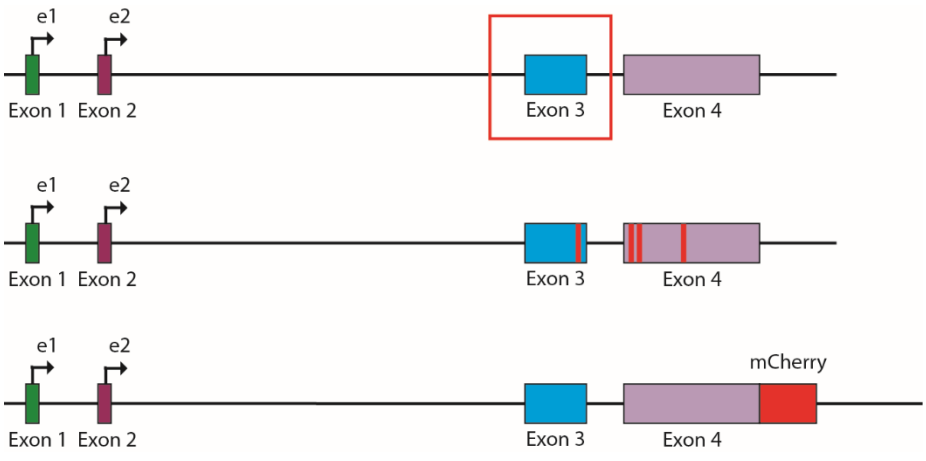


Figure 27. Summary of six CRISPR targeting experiments in LUHMES cells.

MeCP2 KO LUHMES cells were generated by targeting exon 3 of *MECP2*, four Rett syndrome-causing missense mutations were targeted within exons 3 and 4 of *MECP2* and an mCherry tag was added to the 3′ end of exon 4 of *MECP2*.

Experiment	Clone	DNA sequence		RNA sequence	Protein
		allele 1	allele 2		
	H4 (KO2)	14bp del	WT	14bp del	KO
	D10 (KO1)	6bp del	6bp del	Exon 3 removed	KO
1. KO	H5*	1bp ins	11bp ins or WT	1bp ins	KO
sgRNA 16 or 17	C12	1bp ins or WT	7bp del or 6bp del	1bp ins + 7bp del	Practically KO
(puro selection)	F9	2bp ins	27bp del	2bp ins + 27bp del	Reduced
	C11	6bp del	7bp del	WT or 7bp del	Reduced
	A11			WT	Equal
	2_9	R306C	3bp del	R306C	Equal
2. R306C	4_5*	5bp ins	5bp ins	5bp ins	Truncated, equal
sgRNA 15	4_6	WT	35bp del	35bp del	Truncated, reduced
ssODN 1	4_18*	27bp ins	27bp ins	27bp ins	KO
(puro selection)	4_20*			WT	Truncated, v. low
	2_7	3 bp del	8 bp ins	8 bp ins	KO
	2_13	R111G; no RE	R111G; RE	R111G ± RE	Reduced
3. R111G	2_17	R111G; RE	R111G; RE	R111G ± RE	Reduced
sgRNA K2	2_20	R111G; RE; 1 bp ins	R111G; no RE	R111G + 1bp ins	KO
ssODN 4	2_21	R111G; RE; 2 bp sub	R111G; no RE	R111G ± RE	Reduced
	2_22	31bp ins	R111G; no RE	R111G	Reduced
	2_30	R111G; RE	R111G; no RE	R111G ± RE	Reduced
	2_31	R111G; RE	R111G; no RE; 1bp ins	R111G ± RE	Reduced
	4_23	7bp del	12 bp del		
4. R133C	4_28				KO
sgRNA K4	4_37*	19bp del	19bp del		
ssODN 5	4_40	No R133C; RE; mPAM	30bp del		
	4_47	R133C; RE; mPAM	R133C; RE; mPAM	DIED	DIED
5. T158M					
sgRNA K5	5_20	T158M; mPAM	WT	DIED	DIED
ssODN 6					
6. T158M					
sgRNA K6	6_48	6bp del	T158M; mPAM	6bp del	Reduced
ssODN 6					
7. R306C					
sgRNA 5	4		R306C; mPAM; 6bp del	R306C; mPAM; 6bp d	
ssODN 3					
	1G10	R133C; RE; mPAM	R133C; RE; mPAM; 1bp ins	R133C; RE; mPAM	
8. R133C	2E8			24bp del	
sgRNA K4	2D8			GT to AGG	
ssODN 5	2D5			2bp del	
	2E6			7bp ins	
9. R306C	2D4	4bp del	R306C; 18bp del	R306C; 18bp del	
sgRNA 5	2H9			1bp ins	
ssODN 2	1E3			1bp del	

Table 11. Summary Table of all nine CRISPR experiments that have been performed in this project.

DNA sequence, RNA sequence and protein level (if assessed) are detailed for every clone of interest. Asterisks indicate cell lines where more gDNA sequencing should be performed to confirm mutations. All experiments (apart from 1+2) used pX458 and therefore used GFP fluorescence to select for positively transfected cells. Experiments 1+2 used plenti and pX330 respectively and thus puromycin selection was used to select for positively transfected cells.

4.6 Discussion

This chapter has demonstrated the numerous ways in which LUHMES cells can be genetically manipulated in order to create novel cell lines. In particular, for this project, the MeCP2 KO, R306C KI and MeCP2-mCherry tagged cell lines will be studied in more detail.

While preliminary CRISPR experiments were performed in Hek293FT cells, it became apparent that targeting efficiencies of a single sgRNA can vary between these cells and LUHMES, as has been observed for other cell lines (Swiech *et al.*, 2015). The dissimilarities may be due to inherent differences in the ability of the sgRNA to function in the two cell lines, differences in chromatin states or differences in transfection techniques. As repair pathways may be active to different degrees in different cell types (Orii *et al.*, 2006; Shrivastav *et al.*, 2008; van Overbeek *et al.*, 2016), it is perhaps not surprising that efficiencies can vary and emphasises the fact that CRISPR components should be tested in the cell line of interest.

There has been much progress from many different angles towards making Cas9 genome editing more efficient and more precise. Adoption of some of these novel techniques in LUHMES cells will perhaps improve targeting efficiencies, something that would be particularly useful when an experimental design involves knocking in a large tag. As an alternative to transfecting a plasmid that contains the coding sequence for Cas9 and sgRNA, multiple laboratories have demonstrated high levels of INDEL formation and targeted editing by transfection of a pre-assembled Cas9/sgRNA ribonucleoprotein complex (RNP) into cells, as well as *in vivo* editing following injection into mice (Kim *et al.*, 2014; Liang *et al.*, 2015b; Liu *et al.*, 2015; Wang *et al.*, 2016; Zuris *et al.*, 2015). Further to this, Cas9 RNPs seem to induce lower levels of off-target cutting compared to plasmid-expressed Cas9 and sgRNA, probably due to its more restricted period of expression (Kim *et al.*, 2014; Liang *et al.*, 2015b; Zuris *et al.*, 2015). A combination of Cas9 RNPs with cell cycle inhibitors in order to synchronise cells to the same phase of the cell cycle has demonstrated that INDEL and HDR events can be improved (Lin *et al.*, 2014). The specific cell cycle inhibitor that works for LUHMES cells would need to be trialled, as different cell cycle blocking agents worked to different degrees when comparing Hek293T cells, human fibroblasts and hESCs (Lin *et al.*, 2014). Using this approach Lin *et al* managed to improve HDR efficiencies in human fibroblasts and hESCs from undetectable with plasmid transfection to 0.6%, 1.3% and 2% as judged by the T7E1 assay, but the increase in INDEL efficiency was only observed when lower amounts of Cas9 RNP was transfected (Lin *et al.*, 2014). Considering

that high amounts of Cas9 RNP induced equal levels of INDEL formation with or without cell cycle arrest, and did not result in increased off-target mutagenesis, the use of RNPs alone in LUHMES cells may substantially increase editing efficiencies.

As an alternative to transfection of plasmids or Cas9 RNPs, Cas9 has been tagged with a cell-penetrating peptide (CPP) to allow for direct delivery into cells, and CPP:sgRNA complexes have successfully delivered sgRNAs without the need for transfection techniques (Ramakrishna *et al.*, 2014). It is important to note that the Cas9-CPP and sgRNA:CPP components were not complexed into a Cas9 RNP molecule prior to addition to cells, and perhaps this resulted in variable delivery of the two components into individual cells, as population analysis of INDEL efficiencies was fairly low in Hek293T cells (~6-20%). While editing efficiencies were low using this approach, the use of a CPP does allow for the manipulation of hard to transfect cells; although due to the successful transfection of LUHMES cells, perhaps CPPs will not improve efficiencies further. A more relevant approach for LUHMES cells would be to combine RNPs with AAV targeting vectors which have been shown to dramatically increase HDR editing rates (Gaj *et al.*, 2017). Gaj and colleagues knocked in an EGFP-CMV-Puro^{R} cassette into the *nestin* locus in rat C6 glioma cells (neuronal progenitors) using 1kb homology arms. They observed 0.2% EGFP positive cells when using plasmid delivered Cas9 and sgRNA with a plasmid donor molecule, but this increased to 3.7% when using RNPs and an AAV donor molecule. Thus, such an approach could be beneficial for targeted KI of cassettes into LUHMES cells in future targeting experiments.

As an alternative to drugs that block cell cycle progression as described above, drugs that inhibit the NHEJ pathway have been shown to increase HDR rates. Use of Scr7 (a DNA Ligase IV inhibitor) in epithelial cells, melanoma cells, mouse dendritic cells and mouse zygotes increased the rate of HDR events 3-19 fold (Maruyama *et al.*, 2015). While the use of NU7441 (a DNA-PK inhibitor) resulted in a shift of repair pathways from NHEJ to MMEJ in order to introduce large deletions in Hek293T cells (van Overbeek *et al.*, 2016). Once the highest non-toxic concentration of Scr7 for LUMHES cells has been established, it would be easy to apply the drug immediately after Nucleofection in an attempt to shift the balance towards HDR and away from NHEJ in future targeting experiments.

Finally, alternative design strategies for new projects might also help to improve targeting efficiencies, for example use of the PITCh method (Precise Integration into Target

Chromosome) which utilises the MMEJ (microhomology-mediated end joining) pathway (Nakade *et al.*, 2014), or the HITI method (homology independent targeted integration) which utilises the NHEJ pathway (Suzuki *et al.*, 2016) are alternatives to HDR-mediated KI of cassettes. Due to the fact that homology-directed repair is not an efficient repair pathway in many types of cells, both PITCh and HITI utilise alternative, more efficient repair pathways (MMEJ and NHEJ respectively) in order to introduce locus-specific cassettes into cells. While HDR-dependent KI of GFP into the *Tubb3* locus in mouse primary neuronal cultures was approximately 1%, HITI increased this to approximately ~50% by encouraging repair using the NHEJ pathway (Suzuki *et al.*, 2016). Alternatively, HDR frequencies can be improved by intelligent design of the donor molecule. It has been found that having a ssODN that is complementary to the sgRNA and contains all of the mismatched residues downstream of the PAM is the most efficient design strategy for HDR-dependent KI (Richardson *et al.*, 2016). The same lab has recently discovered that ssODN-mediated KI occurs via the Fanconi Anemia DNA repair pathway and does not require Rad51-mediated processes, thus implicating a variety of alternative proteins in the repair process (Richardson *et al.*, 2017) (note, this is published on the bioRxiv preprint server). This knowledge will no doubt lead to drug strategies that target the Fanconi Anemia pathway in order to increase HDR-mediated gene targeting.

In conclusion, CRISPR-based genome editing in LUHMES cells is efficient and using these techniques I have managed to create a variety of modified cell lines, the first example of gene editing in the human pre-neuronal LUHMES cell line. The experiments here have provided insights into the outcomes of CRISPR targeting experiments. For example, ssODN molecules are more efficient than plasmid donors for insertion of point mutations, sgRNA efficiencies can vary between two alternative human cell lines, and Cas9 can target the inactive X chromosome for both NHEJ and HDR repair. Furthermore, insights have been gained about the protein expression level of the R306C and R111G mutant MeCP2 molecules, and progress towards understanding whether in-frame or out-of-frame mutations introduced at different points in MeCP2 will result in instability or complete protein KO have been gained. Future CRISPR work in LUHMES cells could be improved by utilising Cas9 RNP molecules, by altering the length of homology arms on plasmid donor molecules (or indeed removing them altogether as is the case for HITI) and by testing various inhibitors of NHEJ such as Scr7.

Chapter 5

5. Sensitive analysis of MeCP2 expression during LUHMES cell differentiation

5.1 Introduction

There were two purposes for generating the MeCP2-mCherry cell line; first to assess the efficiency of knocking in a large tag into LUHMES cells, and second to generate a fluorescently tagged MeCP2 cell line for functional studies. While the first of these two aims has been discussed in Chapter 4, the second will be discussed in this chapter. By tagging MeCP2 with a fluorescent protein, accurate and sensitive quantifications of MeCP2 protein level in LUHMES-derived neurons during differentiation could be assessed by flow cytometry analysis. Chapter 3.4 described the use of semi-quantitative Western blotting to calculate the molar amount of MeCP2 protein in LUHMES neurons compared to mouse whole brain. While Western blotting is an easy technique, it relies on efficient antibody detection of the antigen of interest, reliable loading controls (for example H3), and is not completely quantitative. Flow cytometry, on the other hand, can measure the amount of a fluorescently tagged protein directly, avoiding the use of antibodies, does not require a loading control, and is more sensitive than Western blot. As such, flow cytometry analysis of the MeCP2-mCherry LUHMES cell line is described in this chapter as a complimentary approach to Western blot analysis (Chapter 3.4).

An alternative experiment that I was keen to perform was to use the mCherry tag as an efficient tool for enriching MeCP2 and its associated proteins from the nuclear lysate of LUHMES-derived neurons. Such a method would be validated by performing Western blots probing for TLB1X and HDAC3, known interaction partners of MeCP2, (Lyst *et al.*, 2013), using the antibodies that have been validated to work with LUHMES cell extracts in this

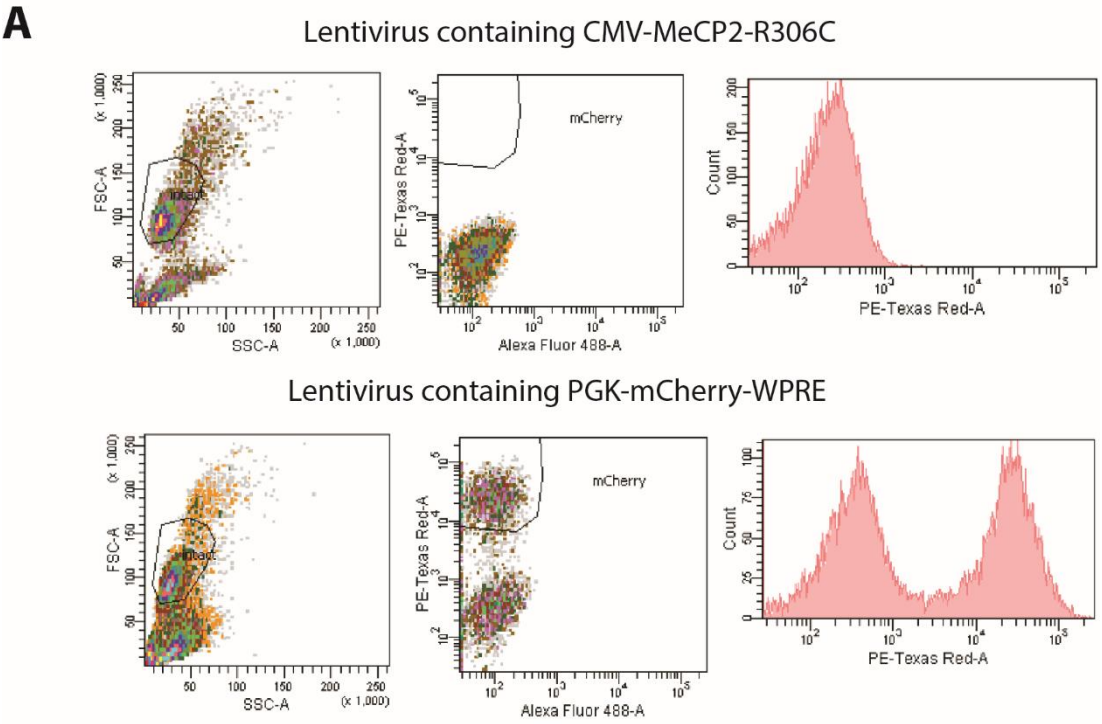
project (Chapter 3.4.2). Upon validation of successful enrichment of MeCP2 and its known interaction partners, samples would be sent to mass spectrometry for sensitive and quantitative analysis. To answer relevant biological questions, comparisons between WT and mutant cells lines, such as R306C, should be performed. In particular, there would be at least two separate hypotheses for an experiment involving an R306C-mCherry tagged cell line. First, interactions with members of the NCoR/SMRT corepressor complex should be abolished or dramatically reduced in the R306C sample relative to the WT sample. Second, interactions in the WT sample that would be suggestive of a link between MeCP2 and RNA regulation (for example Eif4a3, Prpf19 and Snrnp70 which were identified in pull-downs using MeCP2-GFP from mouse brain, performed by Dr Robert Ekiert) would also be abolished in the R306C sample. While this experiment was out-with the scope of this project, assessment of the MeCP2 proteome in human neurons is an interesting and relevant question in order to comprehensively understand the function of MeCP2 in human neurons.

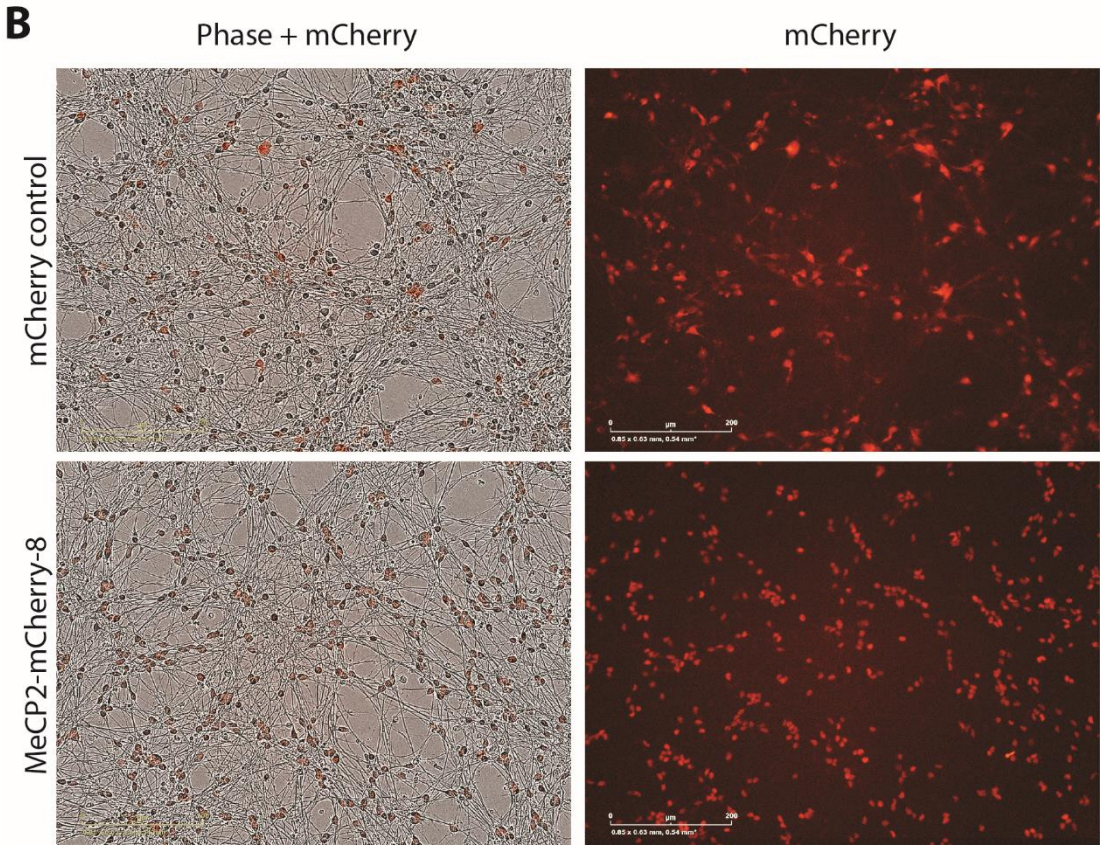
5.2 Generation of an mCherry expressing control cell line

To create an appropriate mCherry-expressing control cell line I generated lentiviruses expressing PGK-mCherry-WPRE using the pLLRSIN.cPPT.PRK.EGFP.WPRE plasmid (gift from Dr Michael Robson, Dr Eric Schirmer's laboratory) and the pBSMe2NSmCherry plasmid (gift from Dr Jacky Guy). The WPRE element is the woodchuck hepatitis virus post-transcriptional regulatory element that increases transgene expression (Higashimoto *et al.*, 2007). Flow cytometry analysis (Figure 28A) and Incucyte fluorescence imaging (Figure 28B) confirmed the generation of LUHMES cells expressing mCherry protein ubiquitously throughout the cell. As can be seen in Figure 28B, mCherry fluorescence in the MeCP2-mCherry-8 cell line is localised to nuclei, whereas mCherry fluorescence in the mCherry control cell line is apparent in cell bodies and in the neurites of LUHMES-derived neurons.

Figure 28. Generation of an mCherry expressing LUHMES cell line.

(A) Flow cytometry analysis of proliferating LUHMES cells demonstrating the presence of PE-Texas Red-A positive cells in the LUHMES population that was infected with lentivirus expressing mCherry. LUHMES cells infected with a virus not containing mCherry (specifically a CMV-MeCP2-R306C lentivirus) are a control for mCherry expression. **(B)** Incucyte images of the mCherry control cell line and MeCP2-mCherry-8 cell line after 9 days of differentiation showing phase contrast pictures of neurite outgrowth and mCherry fluorescence in both cell lines. Scale bar is 200 µm.

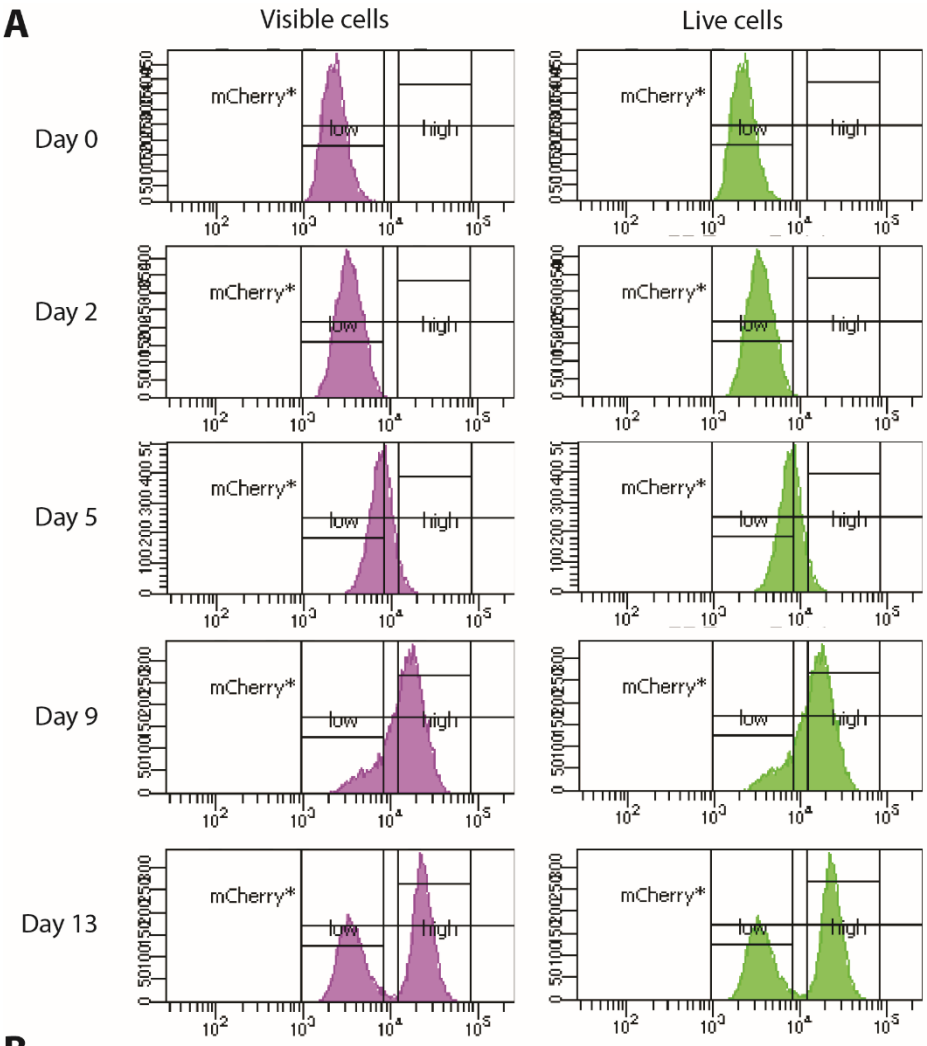

A
Lentivirus containing CMV-MeCP2-R306C
FSC-A
SSC-A
mCherry
PE-Texas Red-A
Alexa Fluor 488-A
Count
Lentivirus containing PGK-mCherry-WPRE


B
Phase + mCherry
mCherry
mCherry control
MeCP2-mCherry-8

5.3 Use of FACS to track MeCP2 expression during differentiation

I used the MeCP2-mCherry tagged cell line in order to sensitively and accurately quantify the levels of MeCP2 protein during LUHMES cell differentiation by flow cytometry. Cells were differentiated for 0, 2, 5, 9 and 13 days and, as expected, during differentiation and as LUHMES-derived neurons matured, the level of mcherry fluorescence increased indicating elevated MeCP2 protein levels (Figure 29). At day 9 of differentiation, a shoulder in the mCherry fluorescence peak was apparent, and by day 13 a 2nd peak in the mCherry fluorescence profile had emerged (Figure 29A). While the mean mCherry fluorescence level in the high mCherry peak continues to increase from day 9 to day 13 as expected (Figure 29B, "high mcherry" row), the appearance of a 2nd population of cells at the latter stages of LUHMES neuronal maturation is surprising.

In order to rule out the possibility that this effect is an artefact of a single clone, I repeated the experiment using a 2nd MeCP2-mCherry cell line (clone 8) and included the mCherry control cell line described in Chapter 5.2. Assessment of mCherry fluorescence levels in these cell lines at days 8, 11 and 14 of differentiation confirmed the presence of the 2nd mCherry-low population in MeCP2-mCherry cell lines (Figure 30A+B). This mCherry-low cell population is not apparent at day 8, the shoulder appears at day 9, a small but distinct population is formed by day 11, and the size of this mCherry-low population increases in abundance by days 13 and 14. The mCherry control cell line also contains two broad populations of fluorescence, but I do not feel that this negates the conclusion that there are two populations in the MeCP2-mCherry cell lines for a number of reasons. First, two populations of fluorescence are present in the mCherry control cell line before differentiation (Figure 28A), indicating that these populations are an intrinsic property of the control cell line. Second, these two populations are present at all stages of differentiation tested (Figure 30A). This indicates that the mCherry high and low populations are not dependent on the maturation of LUHMES-derived neurons, whereas the high and low populations in the MeCP2-mCherry cell lines are dependent on maturation. Third, no clonal selection was performed on this control cell line, and thus the cell population contains a mix of cells that have different numbers of lentiviral genomes integrated and thus would contribute to a broad range of mCherry expression in the population.



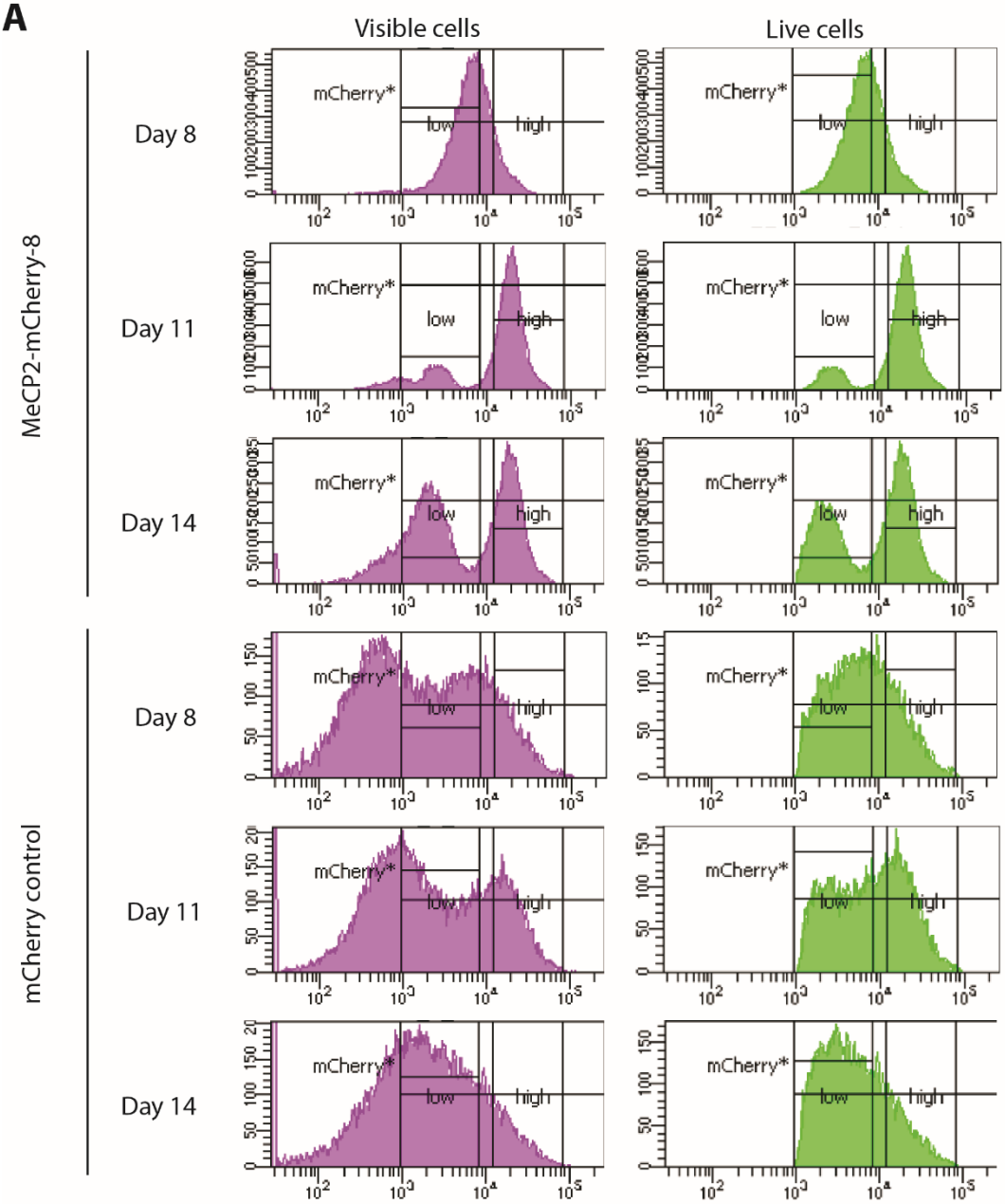
B

Age of cells		Undiff	Day 2	Day 5	Day 9	Day 13
Singlets	# Events	10000	10000	10000	10159	10348
	Mean Fluor	2227	3407	7645	14980	14520
mCherry*	# Events	9936	9987	9997	10148	10330
	Mean Fluor	2238	3411	7647	14995	14544
High mCherry	# Events	8	10	577	6498	5610
	Mean Fluor	18646	17833	14945	19143	23554
Low mCherry	# Events	9917	9938	6521	1860	4531
	Mean Fluor	2216	3374	6192	5204	3582
Live	# Events	9827	9892	9947	9912	10054
	Mean Fluor	2238	3412	7624	15052	14768

Figure 29. Assessment of MeCP2 protein levels during differentiation using MeCP2-mCherry cells.

(A) Flow cytometry analysis of mCherry fluorescence levels in MeCP2-mCherry-7 LUHMES-derived neurons at days 0, 2, 5, 9 and 13 of differentiation. Y-axis is the number of events and X-axis is PE-Texas Red-A fluorescence. **(B)** Quantification of the number of events and mean PE-Texas Red-A fluorescence in the samples shown in (A). Singlets refers to all single cells that were identified by the flow cytometer. mCherry* refers to all cells that were identified as being mCherry positive. Live refers to cells that were identified as being live cells using the LIVE/DEAD Fixable Far-Red Dead Cell Stain (Life Technologies, L10120), see green graphs in (A). High mCherry, Low mCherry and mCherry* gates are shown in (A).

A



B

Age of cells		Day 8	Day 11	Day 14
Singlets	# Events	20055	19321	20021
	Mean Fluor	7441	15016	9634
mCherry*	# Events	19355	17874	17130
	Mean Fluor	7693	16186	11173
High mCherry	# Events	2539	12805	7440
	Mean Fluor	17461	20749	21158
Low mCherry	# Events	12770	3798	8403
	Mean Fluor	5102	2723	2466
Live	# Events	19158	17289	15367
	Mean Fluor	7758	16681	12258

Figure 30. Assessment of MeCP2 protein levels during differentiation using MeCP2-mCherry-8 cells.

(A) Flow cytometry analysis of mCherry fluorescence levels in MeCP2-mCherry-8 LUHMES-derived neurons and mCherry control LUHMES cells at days 8, 11 and 14 of differentiation. Y-axis is the number of events and X-axis is PE-Texas Red-A fluorescence. **(B)** Quantification of number of events and mean PE-Texas Red-A fluorescence in the samples shown in (A). Singlets refers to all single cells that were identified by the flow cytometer. mCherry* refers to all cells that were identified as being mCherry positive. Live refers to cells that were identified as being live cells using the LIVE/DEAD Fixable Far-Red Dead Cell Stain (Life Technologies, L10120), see green graphs in (A). High mCherry, Low mCherry and mCherry* gates are shown in (A).

In order to rule out the possibility that the mCherry low cell population contained were dead or dying neurons that lost fluorescence over time, cells were incubated with the LIVE/DEAD Fixable Far-Red Dead Cell Stain (Life Technologies, L10120). This indicated that a portion of the low-mCherry cells in the control cell line are in fact dead or dying cells (Figure 30A, compare green and purple graphs). The mCherry low population in both MeCP2-mCherry cell lines, however, are not dead or dying cells that are auto-fluorescing, as incubation of the cells with the LIVE/DEAD Cell Stain established that this mCherry low population is part of the live cell population (Figure 29A + Figure 30A, compare purple graphs (all cells) to green graphs (live cells)).

While neurons containing low levels of MeCP2 would not be detected by Western blot analysis, and my immunofluorescence images shown in Figure 4 of Chapter 3.2.2 cannot confirm the presence of an MeCP2-low population of neurons at day 12, the data presented here are indicative of a live cell population in LUHMES-derived neurons that express a decreased amount of MeCP2 protein. The level of fluorescence in the mCherry high population continues to increase during differentiation from an average fluorescence of 14945 at day 5 to 23554 at day 13, indicative of increased expression of MeCP2 (Figure 29B). The average fluorescence of the mCherry low population peaks at a level of 6192 at day 5, and then decreases to a value of 2466 at day 14 as differentiation progresses (Figure 29B + Figure 30B). These values average out to create a fairly constant total mCherry fluorescence value of ~15000 on day 9 onwards (Figure 29B + Figure 30B), in agreement with the result in Figure 7B (Chapter 3.4.1) that the MeCP2 level in LUHMES derived neurons did not increase from day 9 to day 12 of differentiation.

To conclude, MeCP2 expression increases during differentiation until day 9, after which the total MeCP2 amount appears to level out. This total MeCP2 amount is composed of two populations of cells; a low MeCP2 expressing population and a high MeCP2 expressing population. The high MeCP2 expressing population continues to increase the expression of MeCP2 throughout the entire differentiation process. The low MeCP2 expressing population

of cells appears as a separate population of cells at day 11 and while its MeCP2 expression level decreases as differentiation progresses, the number of cells in this population increases.

5.4 MeCP2-mCherry clone summary table

While two MeCP2-mCherry clones have been assessed in detail with respect to MeCP2 protein levels, 28 clones were generated from the targeting experiment. The following table summarises all of the genetic and protein information that has been collected for these clonal cell lines. While the WT PCR assay (primers 1+2) was not designed to amplify the WT allele alone (Supplementary Figure 9A), it became apparent that this PCR was unable to amplify the mCherry allele (Supplementary Figure 9B+C+D). The WT allele is smaller than the mCherry allele thus perhaps there is preferential amplification of this allele in the PCR reaction. Thus the PCR reactions are unable to determine if a cell line is heterozygous or homozygous for the *MECP2-mCherry* allele. Southern blot analysis (performed by Dr Jim Selfridge, see Chapter 5.5) can distinguish between homozygous and heterozygous clones.

5.5 Discussion

The discovery that MeCP2 expression towards the latter stages of LUHMES-derived neuron maturation splits into two peaks with different fluorescence levels is surprising. This effect is not likely to be an abnormal property of a single clone as two MeCP2-mCherry cell lines were assessed. It could be caused by the mCherry tag itself, and thus assessment of WT and MeCP2-mCherry neurons by flow cytometry analysis using an Alexa Fluor 647-conjugated, anti-MeCP2 antibody should be performed (Tillotson *et al.*, under review; adapted for use in LUHMES cells by Laura Fitz-Patrick). The finding that MeCP2 expression is heterogeneous is not novel. Indeed, Balmer *et al* reported a broad peak of MeCP2 expression levels in cerebral cortex samples from adults who died of non-neuronal causes (Balmer *et al.*, 2003). While there are a range of different cell types in the cerebral cortex of the adult human brain, the cortex was chosen for analysis due to its high proportion of neuronal cells. Furthermore, comparison of NeuN+ cell populations in the mouse brain (to identify neuronal cells) with MeCP2 expression by FACS analysis, also found a broad level of MeCP2 expression in the NeuN+ cell population (Skene *et al.*, 2010). Thus the finding that the MeCP2 expression level splits as LUHMES-derived neurons mature could perhaps reflect the *in vivo* situation in the adult human brain. Alternatively, perhaps off-target integration of the mCherry targeting cassette in-frame into a locus that increases its transcriptional output as LUHMES-derived neurons mature could account for the 2nd population of cells.

Clone	PCR1		PCR2		PCR3		PCR4		PCR5		Southern			Western	SP5	Incucyte
	WT	CH	WT	CH	WT	CH	WT	CH	WT	CH	WT	CH	Oth			
3	Y	Y														
6	Y	N			Y	Y	Y	N	Y	Y						
7	Y	Y									Y	Y	8.8	Y	Y	Y
8	Y	Y	Y	Y							Y	Y	8.8	Y	Y	Y
9	Y	Y	Y	Y							Y	Y	8.8			Y
10	Y	Y	Y	Y							N	N	(2)			Y
11	Y	Y														
12	N	Y					Y	Y								
13	N	N					Y	Y	Y	N	Y	Y	8.8			Y
14	N	Y					Y	Y								
15	Y	N							Y	Y+1	Y	N	1	Y		
17	Y	Y														
18			Y	Y	N	N	Δ	N	Δ	N						
19			N	Y	N	N	Y	N	Y	Y						
20					Y	Y	Y	Y			Y	Y	8.8			Y
21	Y	Y	Y	Y							Y	N	1			Y
22	Y	N												N		
26	Δ	N							Δ	Y+1						
27	Y	N							Y	N				N		
33	N	N					Y	Y	Y	Y	Y	Y	8.8			Y
34					N	N	Y	Y	Y	Y						
35					Y	Y	Y	Y								
38					N	Y	Y	Y								
39	N	N					N	N	N	Y						
40	Y	Y														
41	Y	N					Y	Y	N	Y						
44	Y	Y														
46	Y	Y														

Table 12. Summary table of all clonal cell lines generated from the mCherry targeting experiment.

Cell lines that are confirmed as being MeCP2-mCherry positive by at least 3 different techniques are highlighted in green. Cell lines that are MeCP2-mCherry positive as determined by PCR analysis alone, or are positive by 1 assay and questionable by another are highlighted in orange. "WT" subheading indicates either a PCR reaction that will amplify the genomic locus regardless of KI status, or presence of the wild-type band in Southern blot analysis. "CH" indicates either a PCR reaction specific for the mCherry allele or presence of the knock-in allele by Southern blot analysis. "Oth" indicates the presence of other, unexplained bands in the Southern blot analysis, numbers indicating the size of the band in kb. (2) indicates that two extra bands are present in Southern blot analysis. For Southern blot details, see Chapter 5.5. PCR gels are shown in

A more probable explanation, however, is that it is a tissue culture artefact. More analysis into the nature of this mcherry low population of cells might be insightful; for example, by staining for NeuN, doublecortin and other markers of mature and immature neurons one could perhaps discover if there are any intrinsic properties of these cells that differentiate them from the mCherry high population. An MeCP2-mCherry mouse model has recently

been produced by Dr Jim Selfridge (personal communication). Direct comparison of NeuN+ cell populations from different brain regions from this mouse, with MeCP2-mCherry LUHMES-derived neurons at different stages of differentiation would be useful in providing an accurate, sensitive and direct (i.e. no antibodies) comparison of MeCP2 levels between these two systems.

There are a number of potential uses for the MeCP2-mCherry cell line, for example as a screening cell line for therapeutics that reactivate the inactive *MECP2* allele on the inactive X chromosome in human neurons. Briefly, such a project would require identification of heterozygous MeCP2-mCherry clones, targeting of the 2nd allele with an alternative fluorescent reporter such as GFP, application of drugs or CRISPR transcriptional activation molecules, and identification of double fluorescent (i.e. yellow) neurons. In order to assess which of my MeCP2-mCherry clones are heterozygous Dr Jim Selfridge performed Southern blot analysis probing for exon 4 of *MECP2* (Supplementary Figure 7A). While a number of clones were heterozygous for *MECP2-mCherry*, in addition to the WT allele and the KI allele, a 3rd band was present in all cell lines that contained the *MECP2-mCherry* allele (6 cell lines in total, Supplementary Figure 7B). There were two possibilities for what this 3rd band could be. WT AAV genomes are known to integrate into the *AAVS1* locus on chromosome 19 in the human genome, thus perhaps the 3rd band indicated a single-copy integrated plasmid (Kotin *et al.*, 1992; Kotin *et al.*, 1990; Philpott *et al.*, 2002; R.J.Samulski *et al.*, 1991). Alternatively, the circular targeting plasmid could exist as an episome and be propagated through LUHMES cell division. AAV viruses have been reported to exist as episomes in mice, non-human primate and human tissues (Nakai *et al.*, 2001; Penaud-Budloo *et al.*, 2008; Schnepp *et al.*, 2003; Schnepp *et al.*, 2005). The fact that the 3rd band is of a strikingly similar size as the targeting vector, that the presence of the AAV rep protein (a DNA helicase) is required for locus-specific virus integration (Deyle & Russell, 2009; McCarty *et al.*, 2004), and that the observed frequency of integration of rAAV genomes is approximately 0.1% (McCarty *et al.*, 2004) suggests that the plasmid is more likely to exist as an episome. Yet, the single-copy nature of the 3rd band in all cell lines is suggestive of a single, integrated copy of the plasmid. In either case, integration and episomal forms have been reported for the single-stranded genome of rAAV viruses, but the ability of a double-stranded plasmid that contains the AAV ITRs but has not been packaged into AAV molecules has not been reported before.

In order to investigate this further, the Southern blot could be stripped and re-probed for mCherry and plasmid backbone sequences; integration of AAV vectors is thought to be

mediated via the ITRs, and thus only the sequence contained within these elements (i.e. the *MECP2* targeting cassette) should be integrated. Additionally, micrococcal nuclease digestion of chromatin prior to gel analysis using radiolabelled probes could be used to assess the chromatin status of the AAV plasmid (Penaud-Budloo *et al.*, 2008). Finally, sequencing using primers along the length of the plasmid could be performed in order to retrieve a plasmid-DNA junction. If these experiments determine the 3rd band to be an integrated copy of the targeting plasmid, then it is exciting to know that highly efficient, site-specific integration of a cassette can be achieved in LUHMES cells. It would be interesting to determine if this integration also occurred in cell lines that did not undergo correctly targeted KI. If this is indeed the case, future experiments using this pAAV targeting vector (with or without a Cas9 directed towards the AAVS1 locus) could be used for efficient integration of expression cassettes into LUHMES cells.

Regardless of whether this 3rd band is indicative of an integrated or an episomal plasmid, it would be necessary to determine if the MeCP2-mCherry sequence is expressed, in addition to the correctly targeted MeCP2-mCherry KI allele. As the targeting cassette contains no mammalian promoters it would be unlikely for expression to occur. Personal communication with Dr Christina McClure (Prof Giles Hardingham's laboratory, who kindly gave me the Gateway cloning-compatible pAAV vector) has revealed the presence of a putative promoter in the backbone of this plasmid. Analysis of the sequence following this putative promoter has revealed that a stop codon exists between the first ATG downstream of the promoter and the beginning of exon 3 of *MECP2* thus suggesting that if any transcription were to occur, it would only produce a short, 102 bp transcript (Supplementary Figure 7C). In support of this hypothesis, Figure 22B in Chapter 4.3.2 shows that LUMHES cells that were transfected with the mCherry targeting plasmid alone do not exhibit any mCherry fluorescence as assessed by flow cytometry analysis. Furthermore, LUHMES cells that were transfected with an alternative pAAV plasmid which contains a PGK-puromycin^{R}-WPRE cassette instead of the *MECP2-mCherry* targeting cassette, do not retain resistance to puromycin after 3 passages, 1.5 weeks of culture and freeze-thawing (
Supplementary Figure 8). This suggests that the pAAV plasmid, whatever its state in LUHMES cells, is not expressed. In contrast to this, MeCP2-mCherry cell lines 10, 15 and 21 which do not have a correctly targeted allele by Southern blot analysis (Supplementary Figure 7B), are mCherry positive by Incucyte or Western blot analysis (Table 12). This could suggest that the high molecular weight bands found in Southern blot analysis of these cell lines correspond to random integrated plasmid or episomal plasmid that is expressed.

Alternatively, these bands may represent correctly targeted *MECP2-mCherry* alleles that have undergone further genomic rearrangements to disrupt the BamH1 site that was used in the Southern blot analysis. Indeed, PCR analysis of cell line 15 produces a large product that could correspond to the ~7kb Southern blot band (Supplementary Figure 9D). A hypothetical genetic structure of this allele is schematically illustrated in Supplementary Figure 9D, which could explain the PCR and Southern blot band patterns as well as the positive MeCP2-mCherry protein expression observed by Western blot (Figure 22E, Chapter 4.3.2). Similar genetic rearrangements may also occur in cell lines 10 and 21 to account for the positive mCherry expression but the lack of a clean *MECP2-mCherry* allele.

Aside from being used to assess the MeCP2 proteome and potentially being used for an inactive *MECP2* reactivation project, there are a number of other uses for the MeCP2-mCherry cell line. One potential therapeutic angle for treating MeCP2 duplication syndrome would be to use CRISPR targeting in order to KO one *MECP2* allele in each cell. While the MeCP2-mCherry cells are not a model system for studying MeCP2 duplication syndrome, the fluorescent reporter that the *MECP2* allele provides would be a useful tool to assess the efficiency of targeted KO in a population of human neurons using virally-delivered CRISPR components. In line with this, if a stop codon was inserted into the *MECP2* gene locus, drugs that allow read-through of the stop codon (and therefore cause an increase in mCherry fluorescence) could be screened for efficacy as therapeutics for Rett syndrome cases which are caused by nonsense mutations. If the mCherry targeting plasmid is found to exist in an episomal form, these cell lines could be used to directly compare CRISPR targeting efficiencies of plasmid and chromosomal DNA in the same cellular environment.

In conclusion, I successfully managed to generate MeCP2-mcherry tagged cell lines, along with an appropriate mCherry-expressing control cell line. Analysis of these cell lines has revealed the surprising result that at the latter stages of LUHMES cell maturation, two populations of neurons exist that are separated based on their level of MeCP2 protein expression. Further flow cytometry analysis of WT neurons will need to be performed in order to confirm this. While the presence of an integrated or episomal vector in the MeCP2-mCherry cell lines is possible, the lack of expression from this vector indicates that these cell lines will still be useful tools for future projects. Never-the-less future targeting experiments in LUHMES cells should make use of a standard targeting plasmid rather than the pAAV plasmid.

Chapter 6

6. Modelling the functional role of MeCP2 in human neurons

6.1 Introduction

In order to understand the cellular and molecular consequences of loss or mutation of MeCP2 on human neuron homeostasis, a number of assays were performed using the MeCP2 null and R306C LUHMES cell lines. These include morphological assessments such as of the ability of mutant neurons to differentiate into mature neurons, as well as transcriptional analysis of mutant cell lines by RNA-sequencing analysis. While a number of defects have been reported before for mouse and human mutant neurons, such as decreased nuclear volume and shorter and less complex neurites (see Supplementary Tables 1 - 4), these analyses have often been performed on neurons from adult mouse brain, or in tissue culture systems where neurons have been differentiated for a number of weeks. The data here will be the first time such assessments have been performed on young neurons that are only 9-days old, and even on neuronal precursors. As a result, the primary and secondary effects of MeCP2 loss or mutation on neuronal function may be detected.

Throughout this chapter, six different cell lines will be referred to in experiments; WT, E10, F6, KO1, KO2 and R306C (Supplementary Figure 10A). WT denotes the pool of wild-type LUHMES cells from which I perform the targeting experiments. E10 and F6 are two clonal cell lines that have a WT genotype. E10 was derived from the same targeting experiment as the KO clones, and therefore experienced the Nucleofection, puromycin selection and single cell cloning procedures at the same time as the KO clones. E10 was referred to as WTC in Chapter 4.3. F6 was derived from single cell sorting of WT cells and therefore has not

experienced any Nucleofection or puromycin selection conditions. KO1 and KO2 are as described in Chapter 4.3.2 and Figure 19, both are CRISPR-targeted cell lines and while KO2 has a complete KO of MeCP2 protein, KO1 has a small amount of truncated MeCP2 protein remaining (Supplementary Figure 4B). The genetic alteration of KO1 (removal of exon 3) is the same as a published *MeCP2*-null mouse model, which also appears to display low levels of a truncated protein by Western blot analysis and is phenotypically similar to true *MeCP2*-null mice (Chen *et al.*, 2001). Thus cell line KO1 is referred to as a MeCP2 null cell line throughout this project. R306C refers to the cell line that contains the Rett syndrome-causing point mutation of arginine at position 306 to cysteine, which is described in Chapter 4.3.2.

6.2 Morphological analysis of *MECP2* mutant cell lines

6.2.1 Phase contrast and immunofluorescence microscopy confirms neuronal differentiation of *MECP2* KO cell lines

While normal neuronal differentiation is observed in mouse models (Chen *et al.*, 2001; Guy *et al.*, 2001; Moretti *et al.*, 2006) and in most tissue culture neuronal differentiation models (Farra *et al.*, 2012; Li *et al.*, 2013c; Marchetto *et al.*, 2010; Yazdani *et al.*, 2012), one study found reduced levels of TuJ (βIII-tubulin) in neurons derived from RTT patient iPSCs indicating reduced neuronal differentiation potential (Kim *et al.*, 2011). In order to assess the ability of MeCP2-null LUHMES neuronal precursors to differentiate into neurons, phase contrast images following the time course of differentiation were taken. There were no gross morphological alterations observed between two WT clones and two KO clones at day 0, 2, 5 or 9 of differentiation (Figure 31A). In line with this, expression of neurofilament (NF) at day 9 of differentiation was unaltered between all four cell lines (Figure 31B). Staining for MeCP2 using a C-terminal antibody that recognises the truncated protein in KO1 confirms a complete KO of MeCP2 in KO2, and low levels of protein in KO1 (Figure 31B). Further immunofluorescence analysis of NeuN and MAP2 confirms efficient differentiation of MeCP2 KO neuronal precursors into mature neurons (Figure 32A), and analysis of DAT and TH confirms differentiation towards a dopaminergic lineage (Figure 32B). It is interesting to note how DAT and TH staining do not seem to co-localise in the same neurons (Figure 32B). While TH (tyrosine hydroxylase) is a cytoplasmic enzymatic protein in the dopamine synthesis pathway, and DAT (the dopamine transporter) is localised to the cellular membrane, both proteins are markers for dopaminergic neurons and are expected to be

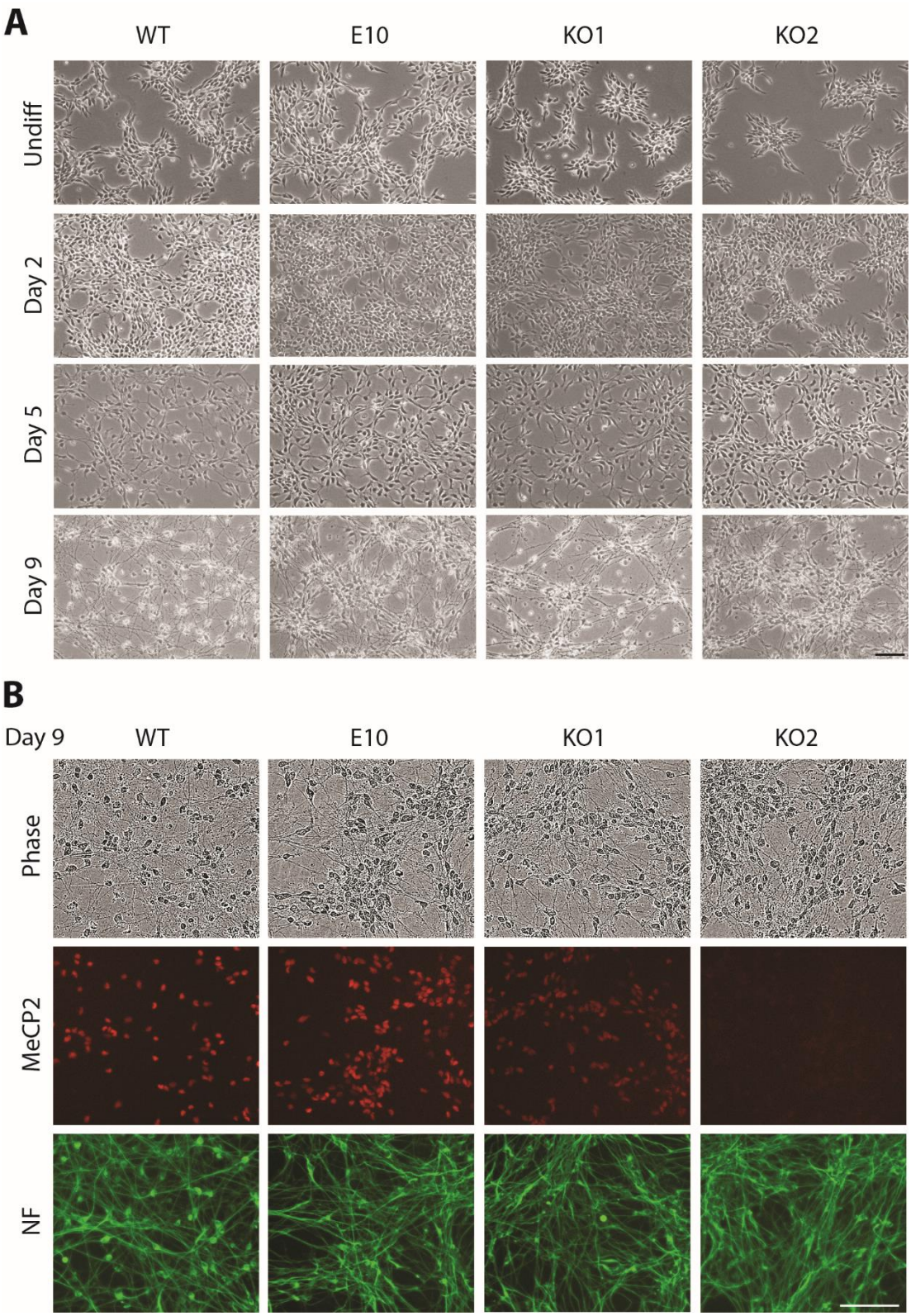


Figure 31. MeCP2 knock-out LUHMES cells differentiate into neurons.

(A) Phase contrast images of cells at different stages of neuronal differentiation. **(B)** Phase contrast and IF images taken by the Incucyte machine of 9 day old neurons showing MeCP2 (red) and neurofilament (green). Scale bars are 100 µm. MeCP2 antibody is D4F3 (Cell Signalling).

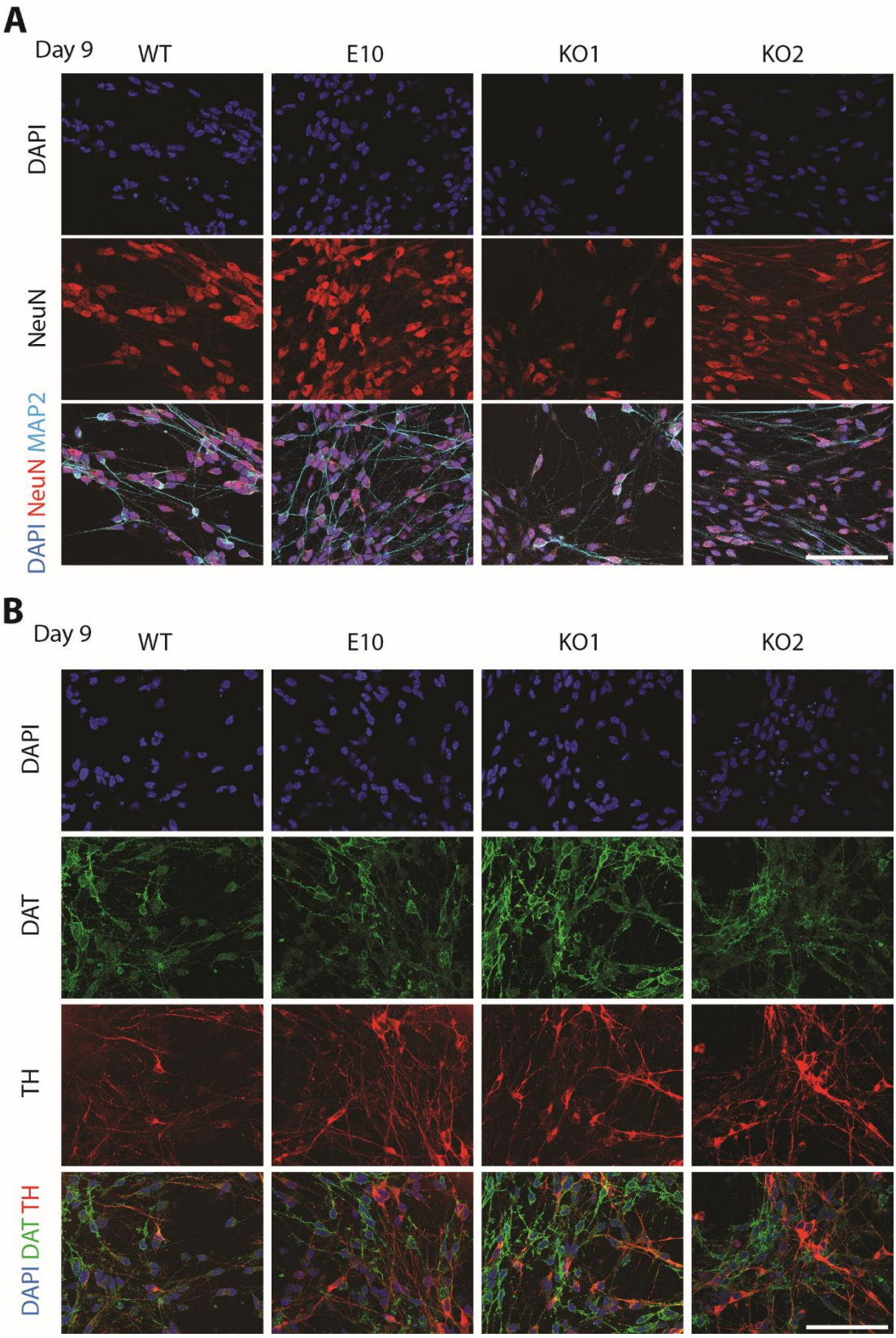


Figure 32. MeCP2 knock-out LUHMES cells differentiate into mature dopaminergic neurons.

(A) IF images of 9-day old neurons showing DAPI (blue), NeuN (red) and MAP2 (cyan). **(B)** IF images of 9-day old neurons showing DAPI (blue), DAT (green) and TH (red). Scale bars are 100 µm.

expressed simultaneously in the same neurons (Liu *et al.*, 2016). Further analysis of this observation should be performed using higher magnifications and alternative antibodies. Overall, these phase contrast and immunofluorescence analyses fail to detect any abnormal differentiation dynamics or a lack of differentiation potential in *MECP2*-null, human neuronal precursors.

6.2.2 Quantitative PCR of 22 different loci in *MECP2* KO neurons demonstrates differentiation dynamics similar to WT cells

In order to assess in a more quantitative manner the ability of MeCP2 KO neuronal progenitors to differentiate into mature neurons, I performed RT-qPCR analysis on 22 different loci at days 0, 2 and 9 of differentiation (Figure 33 + Figure 34). Overall there are no differences between WT and KO progenitors or neurons in the expression level of these mRNAs. Markers of proliferating progenitor cells decrease in all cell lines as differentiation progresses (Figure 33A) and markers of mature neurons (Figure 33B+C), synaptic proteins (Figure 34A) and dopaminergic neurons (Figure 34B) increase. There are some instances where KO2 has an alternative expression level of an mRNA compared to the other three cell lines, for example *SOX2* on day 0, *TUBB3* on day 2 and *NEUROD1* on day 2. As KO1 has the same expression level as the two control cell lines, these dissimilarities are probably due to clonal differences rather than due to the KO of MeCP2. Nevertheless the decrease in *TUBB3* mRNA might be worth analysing in further detail using more KO cell lines, due to the published report that the protein this mRNA produces (TuJ/βIII-tubulin) is reduced in RTT iPSC-derived neurons containing T158M and R306C MeCP2 (Kim *et al.*, 2011). While there are no strong differences between WT and KO cell lines, there are some differences in mRNA expression level between the cloned cell lines (E10, KO1 and KO2) and the uncloned WT cells, suggesting that single cell isolation strategies alter the expression profile of LUHMES cells. Examples include *NDNF*, *SEMA4D* and *SYN* at day 9. Even though the LUHMES cell line is itself a sub-clone, perhaps the process of further cloning affects the expression level of certain genes in a subtle manner or perhaps there is heterogeneity in the population. These alterations, however, do not appear to affect the ability of the cells to differentiate into mature neurons. Overall, there are no striking differences between the expression level or expression dynamics of 22 mRNAs between WT and MeCP2 KO cells.

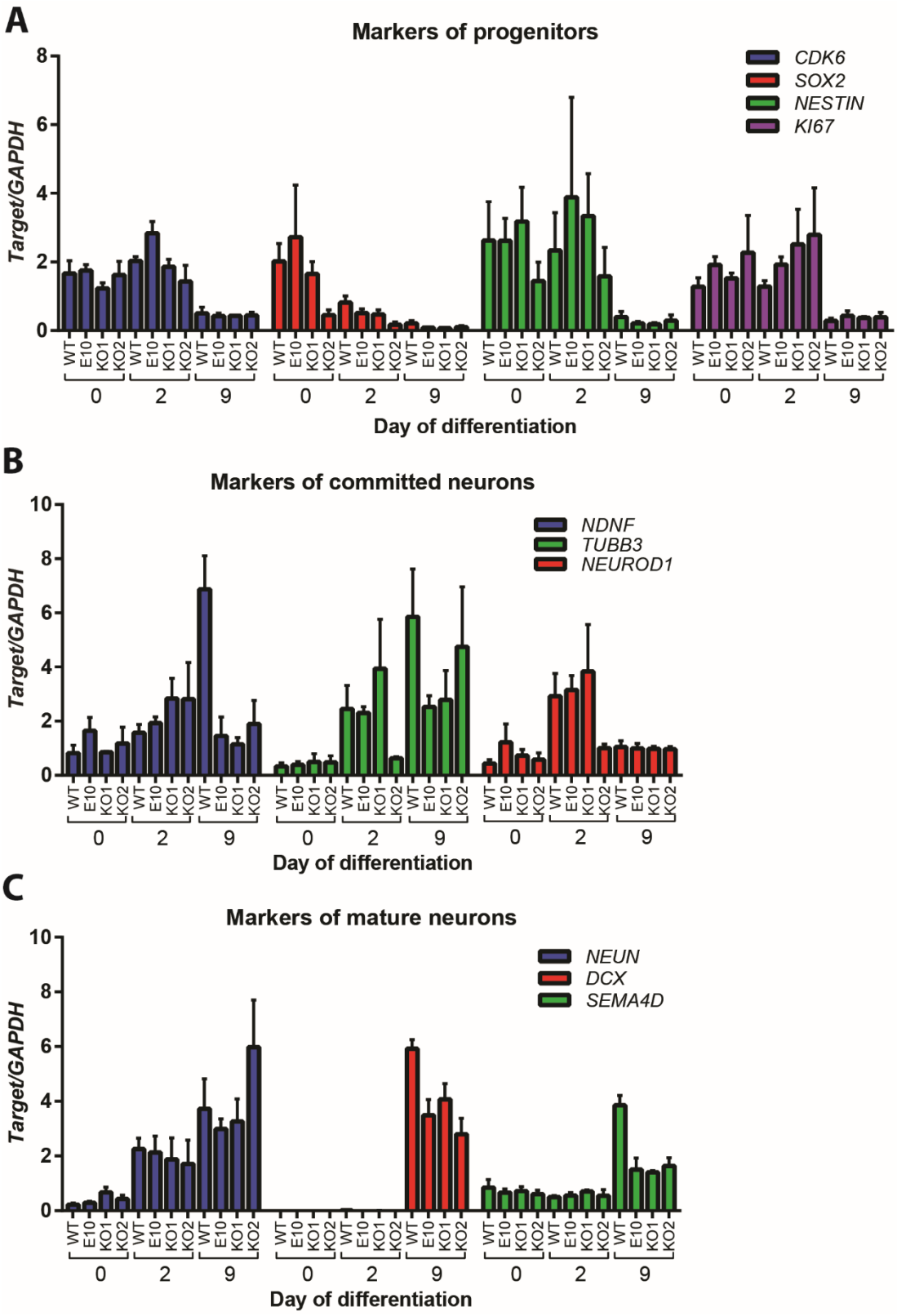


Figure 33. MeCP2 knock-out LUHMES cells differentiate into mature neurons with normal dynamics.

(A) RT-qPCR analysis of four markers of neuronal progenitors. **(B)** RT-qPCR analysis of three markers of committed neurons. **(C)** RT-qPCR analysis of three markers of mature neurons.

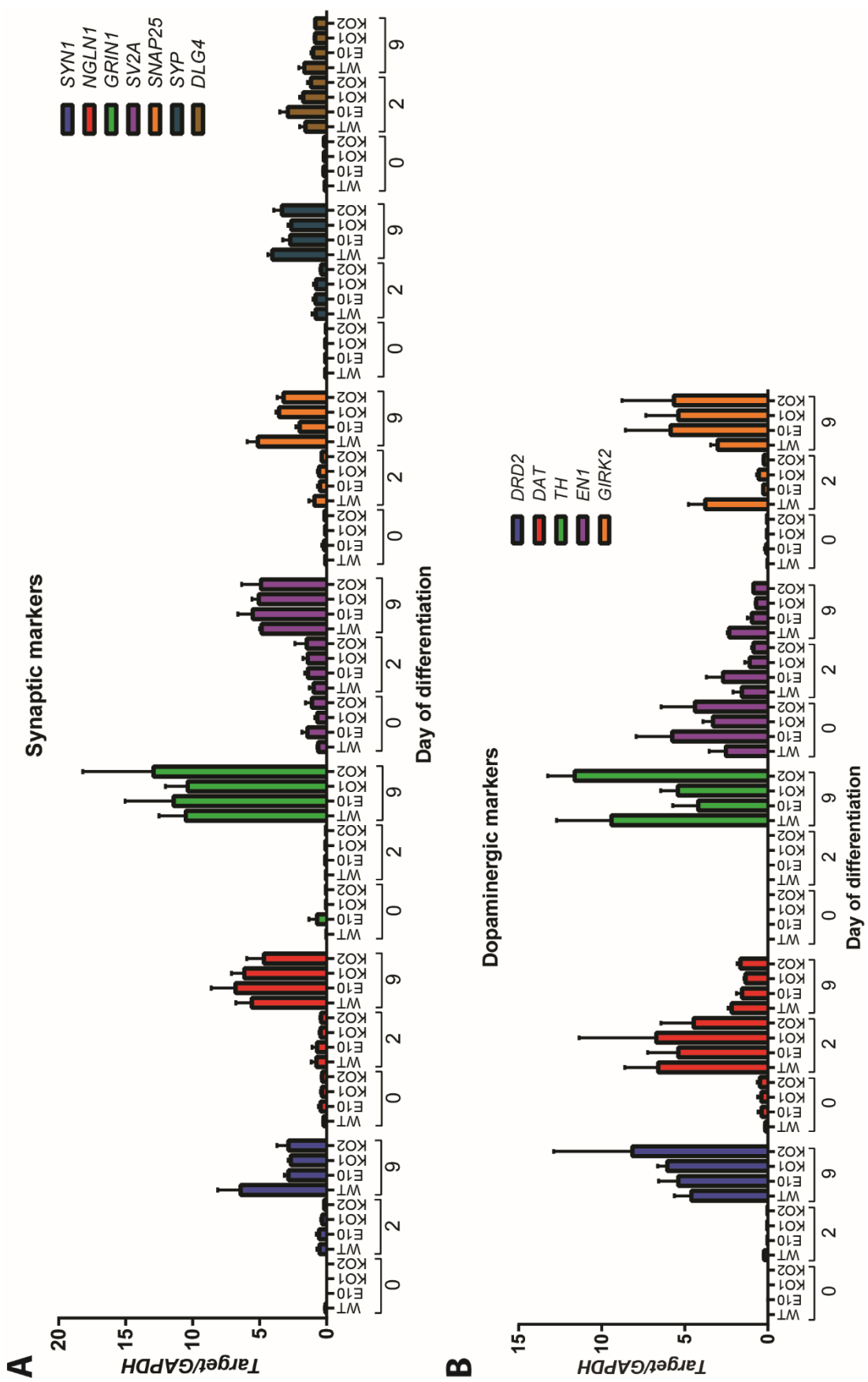

A
Synaptic markers
SYN1
NGLN1
GRIN1
SV2A
SNAP25
SYP
DLG4
Target/GAPDH
Day of differentiation
B
Dopaminergic markers
DRD2
DAT
TH
EN1
GIRK2
Target/GAPDH
Day of differentiation

Figure 34. MeCP2 knock-out LUHMES cells differentiate into neurons that express multiple synaptic and dopaminergic markers.

(A) RT-qPCR analysis of seven mRNA species that produce synaptic proteins. **(B)** RT-qPCR analysis of five dopaminergic markers.

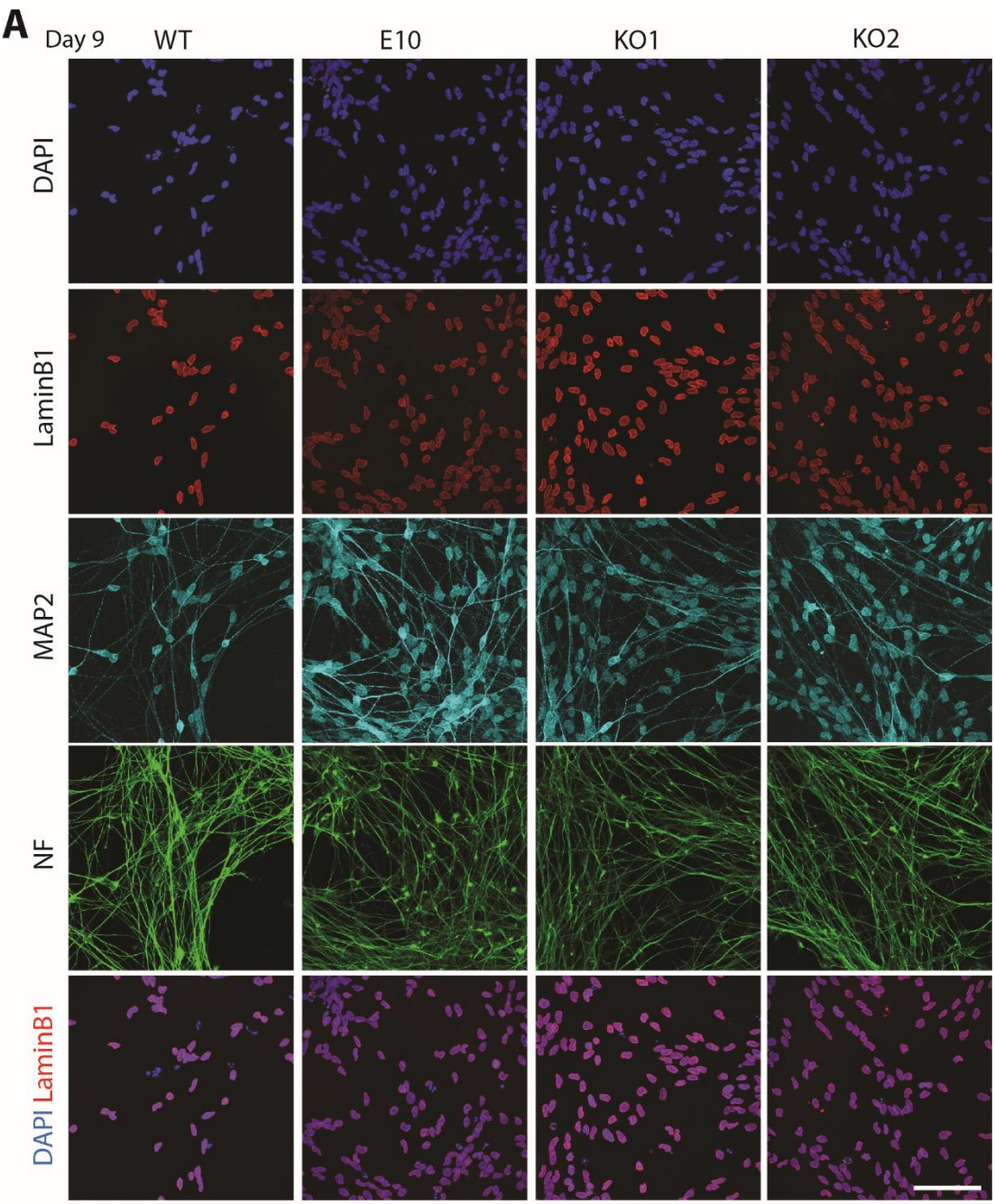
6.2.3 No detectable nuclear volume defects in *MECP2* KO neurons

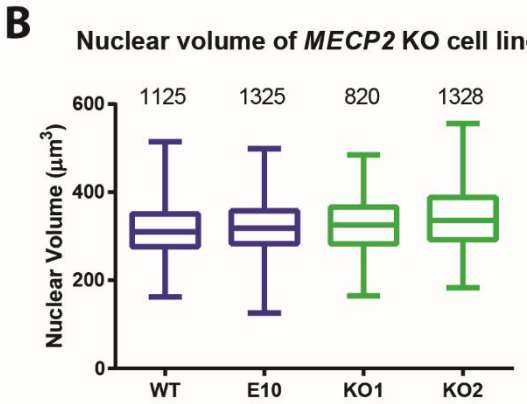
Decreases in nuclear volume have been observed in a number of MeCP2 mutant model systems including RTT patient iPSCs differentiated into neurons and hippocampal slices from *Mecp2*$^{-/y}$ brain (Ananiev *et al.*, 2011; Yazdani *et al.*, 2012). Supplementary Table 3 details these observations, which cover a range of nuclear size decreases from 5% to 33%. Most studies have measured nuclear size (μm^2) using DAPI staining and conventional microscopy, but a recent array tomography approach used DAPI staining of ultra-thin brain slices to measure nuclear volume (μm^3). This study found a 5% decrease in the average nuclear volume of hippocampal CA1, MeCP2 null neurons in female heterozygous *Mecp2*$^{+/-}$ mice (Linhoff *et al.*, 2015). While this result was statistically significant, there was considerable spread in both the WT and null datasets, suggesting a trend towards smaller nuclear volumes in null neurons, rather than an absolute, defined decrease.

In order to assess the nuclear volume of MeCP2 KO LUHMES-derived neurons I differentiated WT, E10, KO1 and KO2 cell lines to day 9 and fixed and stained the neurons with an antibody against laminB1 in order to label the periphery of live nuclei (Figure 35A). The use of laminB1, as opposed to DAPI, removes fragmented DAPI foci from the analysis and therefore removes the confounding effect of dead cells. Furthermore, laminB1 staining gives a sharper and more defined nuclear outline compared to DAPI staining (Figure 35A). Z-stack images were taken on a confocal microscope and stacks were analysed in ImagePro Premier for measurement of nuclear volume. This experiment was performed in triplicate and the cumulative data is shown in Figure 35B. The analysis shows that there is no nuclear volume decrease in MeCP2 KO neurons compared to both WT samples.

Figure 35. No measureable nuclear volume decrease in MeCP2 KO cells.

(A) IF images of 9-day old neurons showing DAPI (blue), LaminB1 (red), MAP2 (cyan) and NF (green). Scale bar is 100 µm. **(B)** Quantification of the nuclear volume by laminB1 staining using ImagePro Premier sofware. Box represents 25-75% of the data with the line representing the median. Whiskers extend to the maximum and minimum values in each dataset. Numbers above each box plot indicate the number of nuclei assessed for each genotye. Nuclei from three biological replicates contribute to each genotype.


A
Day 9
WT
E10
KO1
KO2
DAPI
LaminB1
MAP2
NF
DAPI LaminB1


B
Nuclear volume of MECP2 KO cell lines
1125
1325
820
1328
600
400
200
0
Nuclear Volume (µm³)
WT
E10
KO1
KO2

Perhaps the decrease in nuclear size that has been observed in other systems is a downstream, secondary effect of MeCP2 KO and therefore analysis of older LUHMES-derived neurons may be worthwhile. Alternatively, assessment of neurons using the ChromATin technique might be sensitive enough to detect small alterations in nuclear volume (Linhoff *et al.*, 2015).

6.2.4 Neurite length analysis

In addition to nuclear volume, there have been a number of studies that suggest altered morphology of neurites in MeCP2 mutant samples. Supplementary Table 4 details these observations, which include shorter neurites, decreased dendritic branching and decreased spine densities (Du *et al.*, 2016; Li *et al.*, 2013c). It is important to note, however, that not all studies have found defects in neuronal morphology. Pyramidal neurons from mice that contain a truncated MeCP2 do not display any altered apical or basal dendritic arborisation (Moretti *et al.*, 2006). These mice do display motor abnormalities, social behaviour defects, and memory and learning deficits, suggesting that perhaps some of the phenotypes in RTT model mice are not caused by an underlying neuronal morphology defect (Moretti *et al.*, 2006; Shahbazian *et al.*, 2002b).

As LUHMES-derived neurons do not exhibit a complex neuronal architecture that can be assessed by Scholl analysis (Scholz *et al.*, 2011), neurite length was examined. Neurons were differentiated in 96-well and 6-well TPP plates (Inherent scratches in Nunclon plates affected neurite analysis) and imaged every three hours in an IncuCyte machine in order to observe the growth of neurites during differentiation. The IncuCyte ZOOM software was then trained to distinguish neurites from cell bodies and debris using the NeuroTrack analysis software (Supplementary Figure 11), thus creating a processing definition called "All Older." This processing definition was then assessed for the ability to distinguish neurites in two cell lines during four stages of differentiation (Supplementary Figure 12). While the software accurately determined neurites in the early stages of differentiation, as neurons matured and neurites clumped together, the software was unable to identify these clumps of neurites. Therefore, all analysis performed here was done during the first few days of differentiation.

Assessment of neurite length in two MeCP2 KO and three control cell lines that were differentiated in 6-well plates revealed no difference in neurite length for any KO cell line (Figure 36A). This was confirmed for a second, independent differentiation (Figure 36B) and

for neurites grown in 96-well plates (Figure 37A+B+C). The morphology of MeCP2 KO neurons, and their matched control cell lines, is to be rather more densely packed such that cell body clumps and neurite aggregates form. This phenotype makes it difficult for the IncuCyte software to correctly distinguish a neurite from a cell body (Supplementary Figure 12) and thus could contribute to the observed lack of neurite length phenotype in these cells. Perhaps, co-culture with astrocytes or glial cells will prevent neurite clumping, as has been previously reported (Kuijlaars *et al.*, 2016), and thus re-analysis of MeCP2 KO cell lines under these conditions may be useful.

While a neurite length phenotype is well reported for MeCP2 KO neurons (Supp Table 9.4), neurite defects in neurons that overexpress MeCP2, and thus are a model for MeCP2 duplication syndrome, are less clear cut. A reduction in dendritic complexity has been observed in mouse hippocampal cultures transfected with Flag-tagged MeCP2 (Zhou *et al.*, 2006), and decreased dendrite length and dendritic arbour complexity has been observed upon hMeCP2 overexpression in the developing *Xenopus* (Marshak *et al.*, 2012). On the other hand, overexpression of MeCP2 in transfected mouse cortical cultures resulted in increased length and complexity of both axonal and dendritic processes (Jugloff *et al.*, 2005), and *in vivo* analysis in transgenic mice overexpressing hMeCP2 revealed an increased length, increased number of branch points and increased complexity of dendritic arbours (Jiang *et al.*, 2013a). LUHMES-derived neurons that overexpress MeCP2 have been created by Dr Justyna Cholewa-Waclaw and thus I assessed the neurite length of these using the same protocol as for MeCP2 KO neurons. Overexpression of MeCP2 by ~4-fold and ~12-fold were created by lentivirus infections of *Synapsin* promoter driven *MECP2* and *CMV* promoter driven *MECP2* constructs respectively. Details of the individual cell lines are described in Supplementary Figure 10. Analysis of these cell lines identified a consistent decrease in neurite length for all OE cell lines, compared to all control cell lines (Figure 36C,D,E). A decrease in neurite length was observed at all stages of differentiation, and for three independent differentiations in 6-well plates (Figure 36C,D,E). These results were confirmed by analysis of differentiations in 96-well plates (Figure 37).

To conclude, MeCP2 KO LUHMES-derived neurons do not have a neurite length phenotype, while MeCP2 OE LUHMES-derived neurons exhibit a consistent decrease in neurite length.

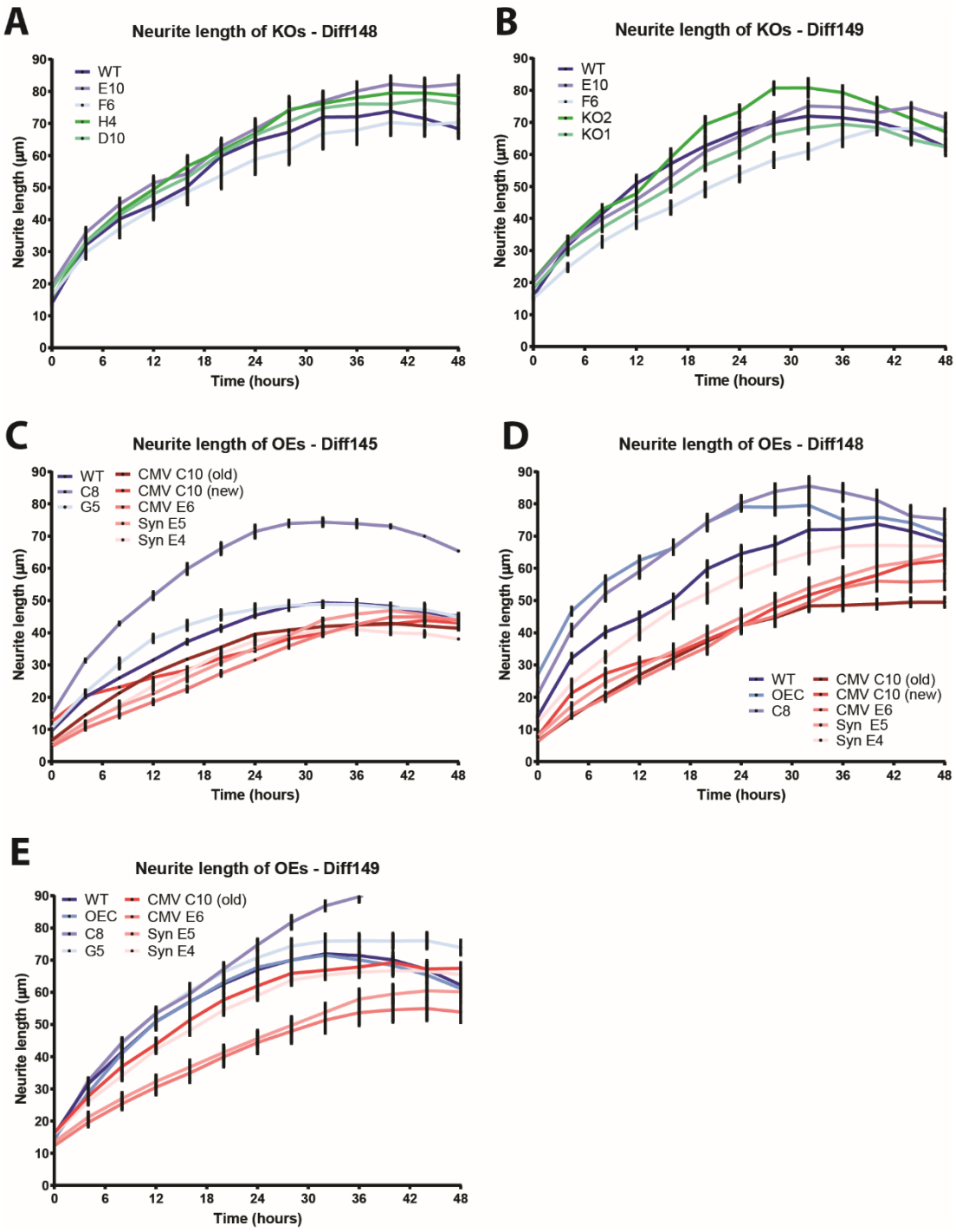


Figure 36. Analysis of neurite length in MeCP2 mutant cell lines cultured in 6-well plates.

(A) Analysis of MeCP2 KO cell lines and their controls for differentiation 148. **(B)** Analysis of MeCP2 KO cell lines and their controls for differentiation 149. **(C)** Analysis of MeCP2 OE cell lines and their controls for differentiation 145. **(D)** Analysis of MeCP2 OE cell lines and their controls for differentiation 148. **(E)** Analysis of MeCP2 OE cell lines and their controls for differentiation 149. All analysis was performed in the IncuCyte ZOOM program using the NeuroTrack software and processing definition All Older. Old and new are two different frozen aliquots of the CMV C10 clone that looked morphologically different upon thawing and therefore needed to be assessed separately for differences in neurite length. Blue lines: WT, green lines: MeCP2 KO, red lines: MeCP2 OE. Time = 0 corresponds to 2-day old neurons.

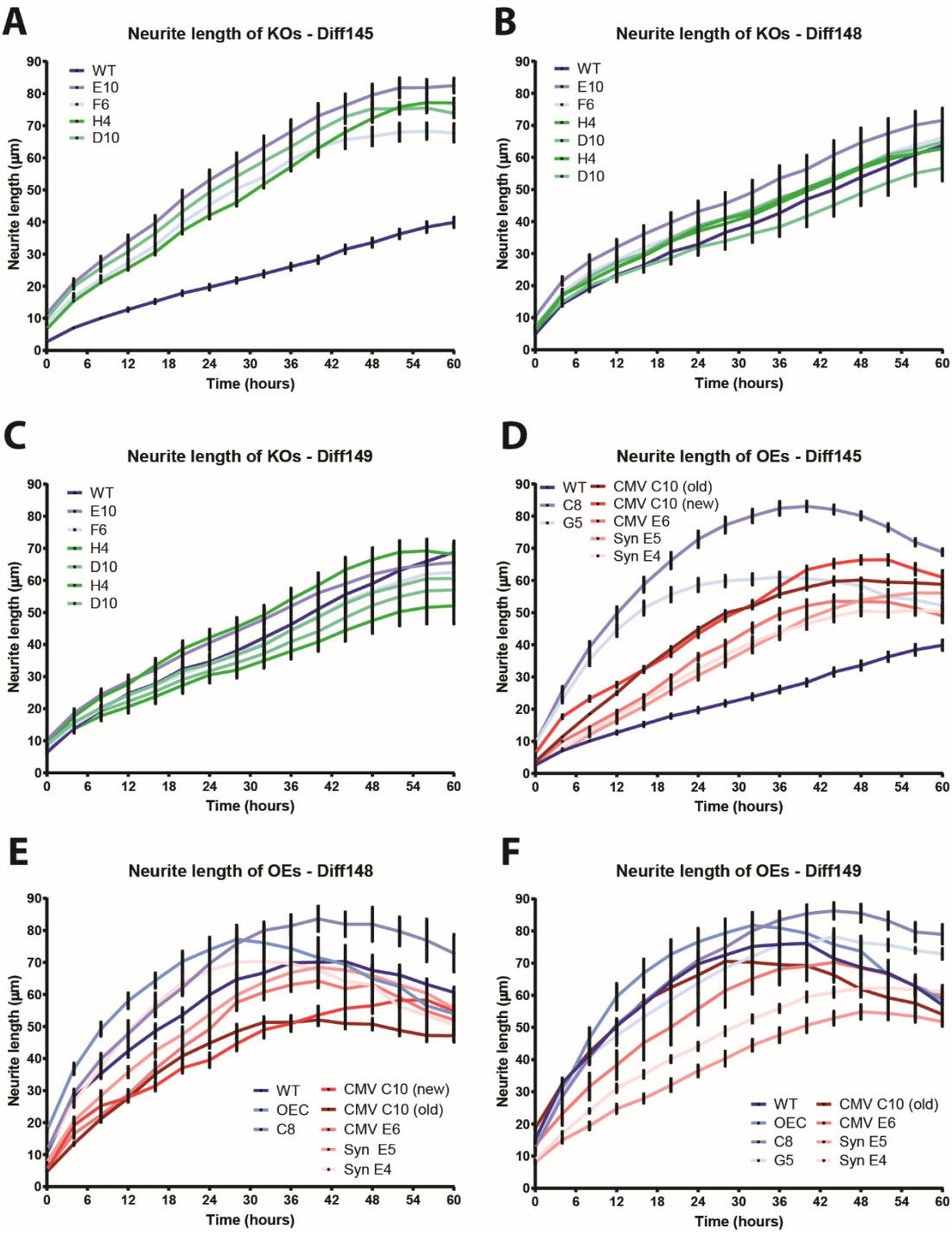


Figure 37. Analysis of neurite length in MeCP2 mutant cell lines cultured in 96-well plates.

(A) Analysis of MeCP2 KO cell lines and their controls for differentiation 145. **(B)** Analysis of MeCP2 KO cell lines and their controls for differentiation 148. **(C)** Analysis of MeCP2 KO cell lines and their controls for differentiation 149. **(D)** Analysis of MeCP2 OE cell lines and their controls for differentiation 145. **(E)** Analysis of MeCP2 OE cell lines and their controls for differentiation 148. **(F)** Analysis of MeCP2 OE cell lines and their controls for differentiation 149. All analysis was performed in the IncuCyte ZOOM program using the NeuroTrack software and processing definition All Older. Old and new are two different frozen aliquots of the CMV C10 clone that looked morphologically different upon thawing and therefore needed to be assessed separately for differences in neurite length. Blue lines: WT, green lines: MeCP2 KO, red lines: MeCP2 OE. Time = 0 corresponds to 2-day old neurons.

6.4 Transcriptional defects in *MECP2* KO and R306C neurons

Due to the ability of MeCP2 to bind to sites of methylated DNA both *in vitro* (Meehan *et al.*, 1992; Nan *et al.*, 1993) and *in vivo* (Klose *et al.*, 2005; Lewis *et al.*, 1992)(Skene *et al.*, 2010), and to repress transcription in *in vitro* repression assays (Nan *et al.*, 1997), it is thought that MeCP2 functions to repress transcription. Evidence in support of this view includes the list of transcriptional corepressors that MeCP2 has been shown to interact with, including Sin3a (Nan *et al.*, 1998), ATRX (Nan *et al.*, 2007) and NCoR/SMRT (Lyst *et al.*, 2013), and recent *in vivo* RNA-sequencing studies (Chen *et al.*, 2015; Gabel *et al.*, 2015; Lagger *et al.*, 2017). As a result, there has been much emphasis to define the transcriptional profiles in MeCP2 mutant models in order to explain disease pathology. Initial experiments focussed on the identification of MeCP2 target loci, whose promoters MeCP2 binds to, and whose expression was altered in MeCP2 null samples. *BDNF* has been identified as a potential MeCP2-regulated gene; MeCP2 was shown to bind to the *BDNF* promoter, and BDNF expression was increased in *MeCP2*$^{-/y}$ cortical neurons (Chen *et al.*, 2003; Martinowich *et al.*, 2003). The observation, however, that *BDNF* and indeed other MeCP2-regulated genes were not consistently misregulated in different transcriptome datasets led the field to move from a target gene-centric view of MeCP2 function, to a global view. Recent genome-wide RNA-sequencing experiments have found large numbers of misregulated genes in MeCP2 KO datasets, but the observation that genes are increased and decreased upon MeCP2 mutation or KO complicates the picture (Chen *et al.*, 2015; Gabel *et al.*, 2015; Lagger *et al.*, 2017; Rube *et al.*, 2016). Despite these common global transcriptional profiles there is still much debate as to whether MeCP2 activates or represses genes, and how it brings about these changes; binding to mCG + mCAC, binding to long genes, binding to regions of GC-rich unmethylated DNA. All of these studies have thus far focused on different regions of the mouse brain, thus analysis of a homogeneous population of human neurons might reveal a clearer mechanism for how MeCP2 regulates transcription.

6.4.1 HPLC analysis of total RNA

By measuring the amount of incorporated radiolabelled [^{32}P]-UTP in isolated nuclei from *MeCP2*$^{-/y}$ mESCs differentiated into neurons, Yazdani and colleagues discovered that mutant neurons had approximately 40% of the transcriptional output of WT neurons (Yazdani *et al.*, 2012). This was the first assessment of total RNA levels in MeCP2 mutant neurons as all previous transcriptional analysis had focussed on steady-state measurements of mRNA by RNA-sequencing or microarray techniques. Similar observations have beens made in hESCs

differentiated into neurons, where MeCP2 KO neurons had 27% and 47% less total RNA compared to isogenic controls at 2 weeks and 4 weeks of differentiation (Li *et al.*, 2013c), and in olfactory neuroepithelium cells that had 25% less RNA (Rube *et al.*, 2016). I therefore measured the total RNA in LUHMES-derived neurons using a highly accurate methodology.

HPLC (high-performance liquid chromatography) has been used to measure the total amount of RNA in regions of the mouse brain by comparing the ratio of ribonucleosides to deoxyribonucleosides in WT and KO samples. Using this approach a ~13% decrease in total RNA in 6-week old *Mecp2*$^{-/y}$ hypothalamus, cortex and cerebellum compared to *Mecp2*$^{+/y}$ controls was found (Lagger *et al.*, 2017) (also shown in Table 13). This technique involves the direct lysis of material in lysis buffer, a proteinase K treatment, and ethanol precipitation of total nucleic acid content from cells. The lack of column purifications during nucleic acid extraction reduces the chance of losing material. Furthermore, the simultaneous deoxyribonucleoside measurement provides an internal control for cell number and thus avoids the use of cell counting machines which have a margin of error.

Using this technique, the total RNA content in MeCP2 KO, OE and R306C LUHMES cells was measured. The total RNA amount was found to be reduced in undifferentiated neuronal progenitors from KO and R306C samples (Figure 38A); a novel finding as RNA amounts have not been assessed in neuronal progenitors before. This defect remained throughout differentiation resulting in a decrease of approximately 9% in 9-day old LUHMES-derived neurons (Figure 38B, Table 13). On the other hand, there is no consistent defect in MeCP2 OE cell lines, in the undifferentiated state or as neurons (Figure 38 C+D). Although in the undifferentiated state, the 12X OE cell lines have a statistically significant decrease in total RNA amount when compared to both controls, the 4X OE does not. This difference could be accounted for by the fact that the *CMV* promoter that drives the 12X OE is active in LUHMES progenitors, whereas the *Synapsin* promoter that drives the 4X OE is not. Analysis of MeCP2 protein level in the undifferentiated cells would be necessary to confirm whether overexpression of MeCP2 in neuronal progenitors causes the decrease of total RNA in the 12X OE cell lines observed here. While overexpression of MeCP2 causes a small but statistically significant decrease in total RNA at day 9 when compared to clonal WT cell lines, there is no change when compared to the lentivirus infected control cell line (Figure 38 XD). The variability in total RNA amount between the controls, in the undifferentiated state and in 9-day old neurons, affects the ability to draw a firm conclusion about the effect of

MeCP2 overexpression in LUHMES-derived neurons on total RNA levels and suggests that any effect of MeCP2 overexpression is small.

As the LUHMES MeCP2 KO data was in agreement with the data from MeCP2 KO mouse brain regions (Table 13), I assessed if a decrease in total RNA could also be observed in R306C KI mice. Previous analysis of KO and OE mice was performed at 6-weeks of age (Lagger *et al.*, 2017). R306C mice, however, are not particularly symptomatic at this age and thus analysis was performed at 11-weeks when they have a similar phenotype as the MeCP2-null mice do at 6 weeks (Brown *et al.*, 2016). Similar to previous HPLC analysis of mouse brain regions, WT cerebellum has a dramatically reduced level of RNA compared to the other WT brain regions (Figure 38E) (Lagger *et al.*, 2017). Figure 38 shows that the R306C point mutation causes a decrease in total RNA in the hypothalamus and cortex of 11-week old mice, but no significant difference was observed in the cerebellum (Table 13).

	LUHMES d0	**LUHMES d9**	**Hypothalamus**	**Cortex**	**Cerebellum**
KO	3.6%	8.8%	*14.9%*	*12.0%*	*10.3%*
OE	9.3%	0.8% ↑	*3.7%*	*3.2%*	*3.5%*
R306C	4.7%	8.7%	12.6%	9.3%	12.0%

Table 13. Percentage decreases in total RNA amount in MeCP2 mutant model systems.

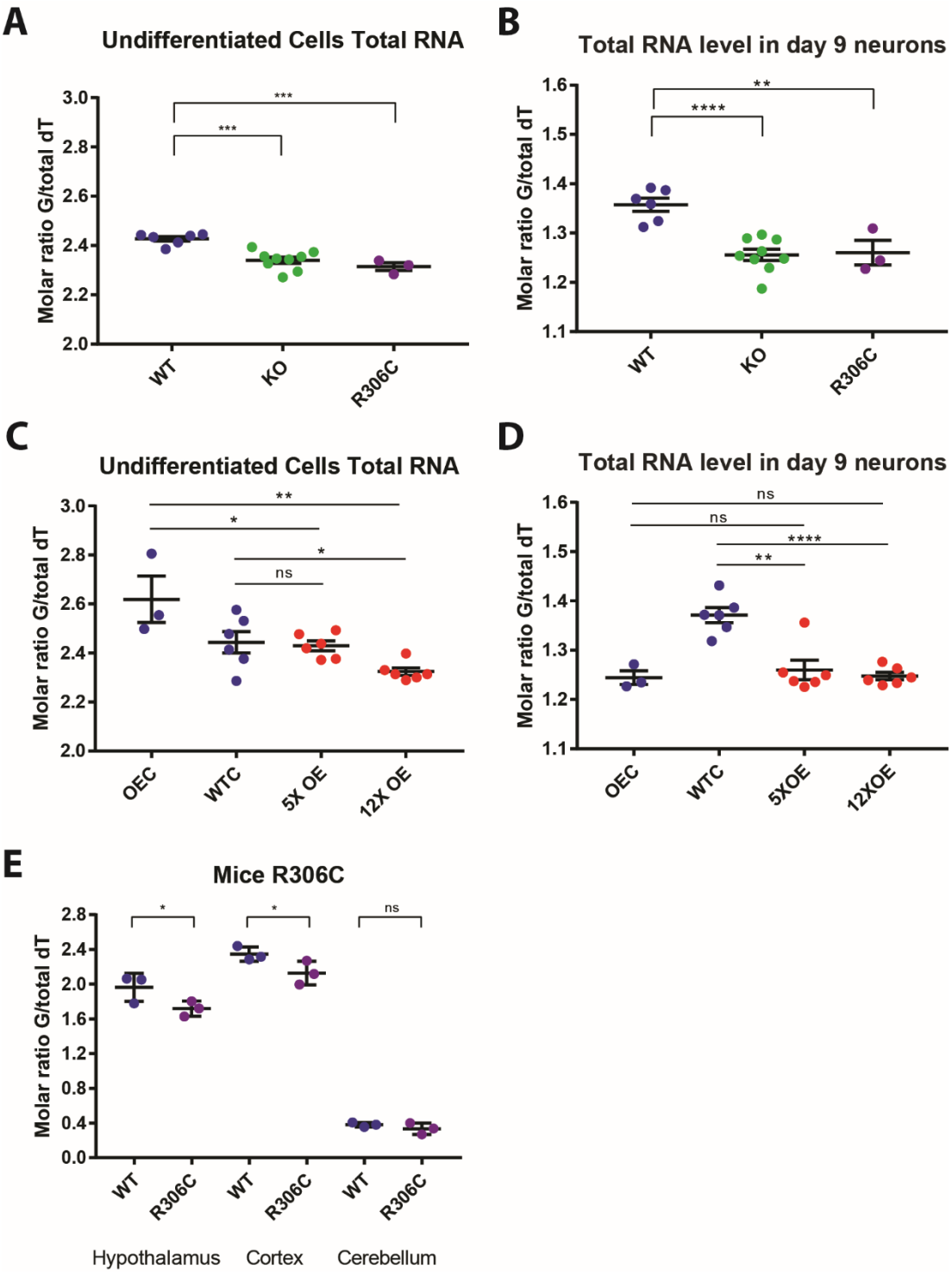
LUHMES OE numbers are the average of 4X and 12X OE samples. Mouse KO and OE samples are from 6-week old mice. Mouse R306C samples are from 11-week old mice. Numbers coloured in green are statistically significant. Data in italics are from (Lagger *et al.*, 2017) where the ratios are G/dC. All other ratios are G/dT. Mouse samples are compared to WT littermates, LUHMES KO and R306C samples are compared to E10 and F6 controls, LUHMES OE samples are compared to the OEC control.

The lack of a strong phenotype in MeCP2 OE mice and LUHMES-derived neurons could be due to the young age of analysis. Mice that overexpress MeCP2 have a mild phenotype at 6-weeks of age, which progresses as the mice age (Collins *et al.*, 2004). Thus analysis of OE samples at later stages of development might reveal more striking differences in total RNA levels. In addition, it may be interesting to analyse pre-symptomatic KO and R306C mice in order to assess whether decreases in RNA amount are a cause or consequence of the RTT-like phenotype in mice.

Figure 38. HPLC analysis of total RNA levels in mutant cell lines and mouse brain regions.

(A) Analysis in undifferentiated KO and R306C cells. **(B)** Analysis in 9-day old KO and R306C neurons. **(C)** Analysis in undifferentiated OE cells. **(D)** Analysis in 9-day old OE neurons. **(E)** Analysis of regions in the mouse brain of 11-week old R306C male mice. For plots A-D, WT indicates a

grouping of E10 and F6 WT cloned controls. KO indicates a grouping of D10, H4 and H5 KO cell lines. OEC indicates the WT cell line that has been infected with a lentivirus that did not contain *MECP2*. WTC indicates a grouping of C8 and G5 WT clones controls. 4X OE indicates a grouping of Syn E4 and Syn E5 cell lines.12X OE indicates a grouping of CMV C10 and CMV E6 cell lines. Each cell line is repeated in triplicate (3 separate differentiations). Statistical significance analysed using an unpaired students two-tailed t-test (or one-tailed for mouse samples) with a cut-off of $p<0.05$.



To conclude, the total level of RNA is decreased in MeCP2 null and MeCP2-R306C proliferating LUHMES progenitors and LUHMES-derived neurons, as well as in brain regions form 11-week old MeCP2-R306C mice. MeCP2 overexpression in LUHMES progenitors may also result in a decrease in total RNA; a decrease that does not remain in LUHMES-derived neurons.

6.4.1 Bulk RNA sequencing analysis

Most evidence points towards MeCP2 playing a role in the negative regulation of gene expression. There have been a large number of studies using RNA-sequencing or microarray techniques to assess the transcriptional changes that occur with MeCP2 mutation or deletion. As MeCP2 is known to bind to methylated cytosines, these gene expression changes are often related to the level of methylation within the gene, i.e. the number of MeCP2 binding sites within the gene.

Dr Justyna Cholewa-Waclaw has produced bisulphite-sequencing, TAB-sequencing (Tet-assisted bisulphite) and oxBS-sequencing datasets from WT LUHMES-derived neurons, allowing the entire methylome to be mapped. This analysis revealed that LUMHES-derived neurons contain very low amounts of mCA, and hydroxymethylation. All transcriptome data in this project was compared to densities of mCG as this modified dinucleotide sequence comprises ~93% of the total methylation in LUHMES-derived neurons. Methylation density/kb was analysed for each gene rather than methylation amount or methylation density, in order to normalise for the length of the gene. This method of analysis stems from the hypothesis that it is the density of methylated cytosines within a gene (i.e. the density of MeCP2 binding sites), rather than the number of methylated cytosines (i.e. the absolute number of MeCP2 binding sites) per gene that is important for MeCP2-mediated regulation of transcription (Figure 39).

RNA-sequencing was performed in quadruplet for two KO cell lines (KO1 and KO2) and two wild-type cell lines (E10 and F6). Venn diagram analysis revealed 5629 significantly changing genes in the KO1 to control comparison, with 5068 significantly changing genes in the KO2 to control comparison (Figure 40A+B). Previous microarray analysis of *Mecp2*-null mouse cerebellum revealed 30% fewer misregulated genes when low levels of a truncated MeCP2 protein was expressed compared to complete KO of all protein (Jordan *et al.*, 2007). This result led to the hypothesis that low levels of a truncated MeCP2 protein results in fewer misregulated genes compared to a complete knock-out of all protein, and thus the

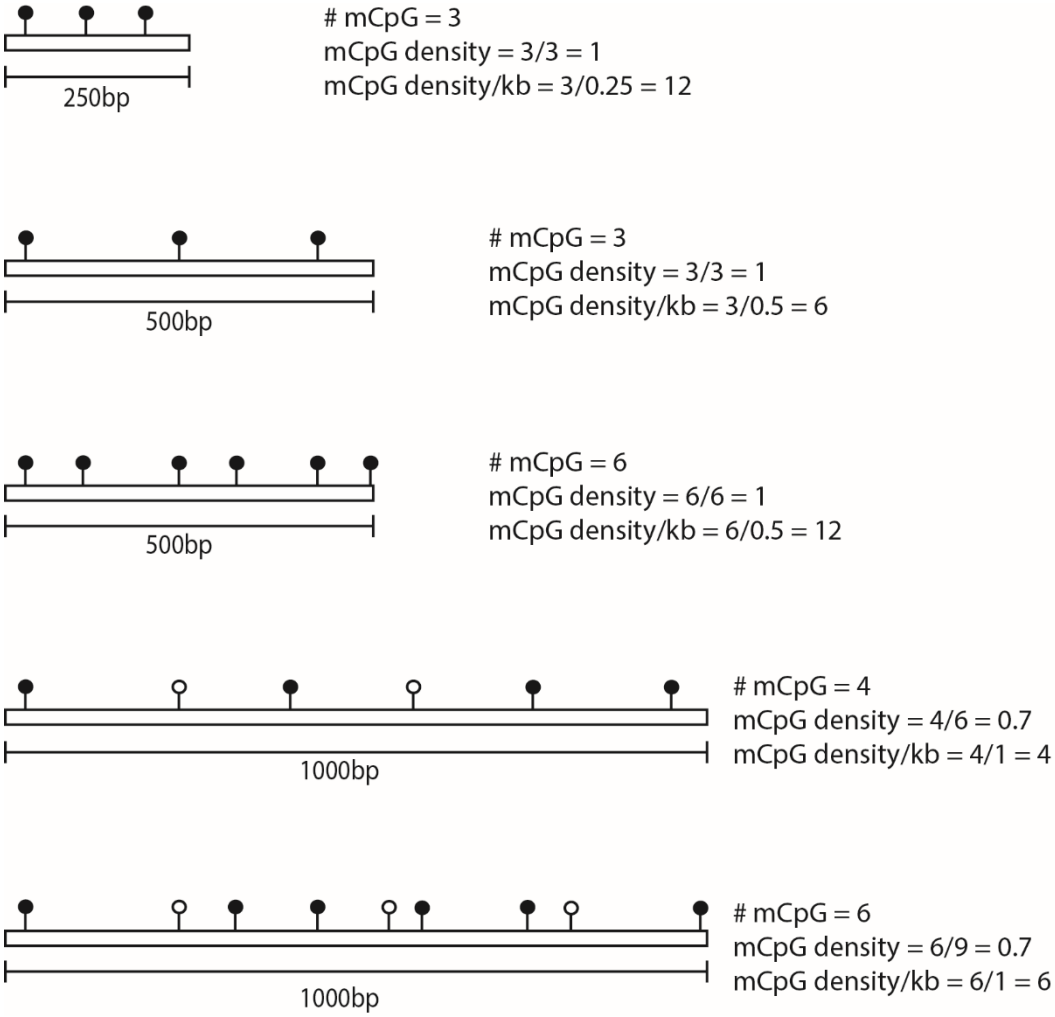


Figure 39. Principles of methylation amount, methylation density and methylation density/kb.

Five genes of different lengths are shown. Each gene has a different number of CpG dinucleotides (lollipops), and different numbers of methylated (black lollipops) and unmethylated (white lollipops) CpGs. The mCpG number, mCpG density and mCpG density/kb is shown for each gene on the right hand side.

truncated protein (which lacks coding from exon 3) may retain partial MeCP2 functionality despite having lost half of its MBD (Jordan *et al.*, 2007). The LUMHES data here contradicts this previous hypothesis as comparable numbers of genes are misregulated in KO1 (low levels of a truncated protein due to loss of coding from exon 3) and KO2 (complete MeCP2 KO). The differences in the number of misregulated genes observed in the previous study could be due to differences in the strain background of the mice rather than a reflection of the functional relevance of a truncated MeCP2 protein (Jordan *et al.*, 2007).

Averaging the datasets for E10 and F6 control cell lines, and comparing this aggregate control dataset to each KO cell line reveals a total number of 3850 significantly misregulated

genes, a third of which are consistent between the two KO datasets (Figure 40C). In order to assess how these genes are changing and what properties might account for the misregulation, the log_2fold-change between KO and WT was plotted against various parameters. First, genes were grouped into eight bins of equal size according to their promoter mCG density (Figure 40D). This analysis revealed no correlation between the direction or scale of a genes misregulation, and the methylation density of its promoter. Plotting instead against gene body mCG density revealed a strong correlation, with genes that contain more gene body methylation being more upregulated in the KO cell lines (Figure 40E). While there is a large spread of data in each of the bins, there is a shift of gene density from below the zero line to above it as mCG density increases, with the average log_2FC for each bin (denoted by white dots) increasing in value with more gene body mCG density. Previous analysis of RNA-sequencing data from mouse identified a gene-length dependent effect on transcription in multiple brain regions whereby longer genes were more upregulated in KO tissues (Gabel *et al.*, 2015). No gene-length dependent upregulation of genes was observed in the LUHMES dataset, and in fact it seemed that longer genes were actually downregulated in KO neurons (Figure 40F). Assessment of the methylation complement of these genes revealed that longer genes tend contain lower amounts of gene body mCG density/kb, and thus the apparent downregulation of genes is likely to be due to the effect of less methylation, rather than due to an intrinsic property of gene length (Figure 40F). The discrepancy between the Gabel dataset and this studies dataset could be explained by the lack of mCA in LUHMES-derived neurons, as the long-gene effect is attributed more strongly to mCA than to mCG (Gabel *et al.*, 2015). Another difference between the two studies is how the methylation analysis is performed. Here, the methylation density of a gene is plotted relative to its total length (mCpG density/kb in Figure 39), while in the Gabel study the methylation density was calculated by summing all methylated cytosine's in the gene and dividing it by the total number of cytosine residues, with no scaling for gene length (mCpG density in Figure 39). This LUHMES project concludes that the denser the gene body methylation, the more upregulated a genes expression will be in the MeCP2 KO neurons. Recent re-analysis of the Gabel data, also came to the same conclusion when gene length was normalised for (Kinde *et al.*, 2016). Finally, analysis of all 13, 154 protein coding genes also revealed a correlation between the fold-change in expression of a gene and the level of gene body mCG density (Figure 40G). The presence of a gene-body methylation- dependent correlation with gene expression changes in the entire transcriptome dataset, that is strengthened when significantly changing genes are selected for, suggests a role for MeCP2 in the repression of gene expression by binding to sites of methylation within gene bodies.

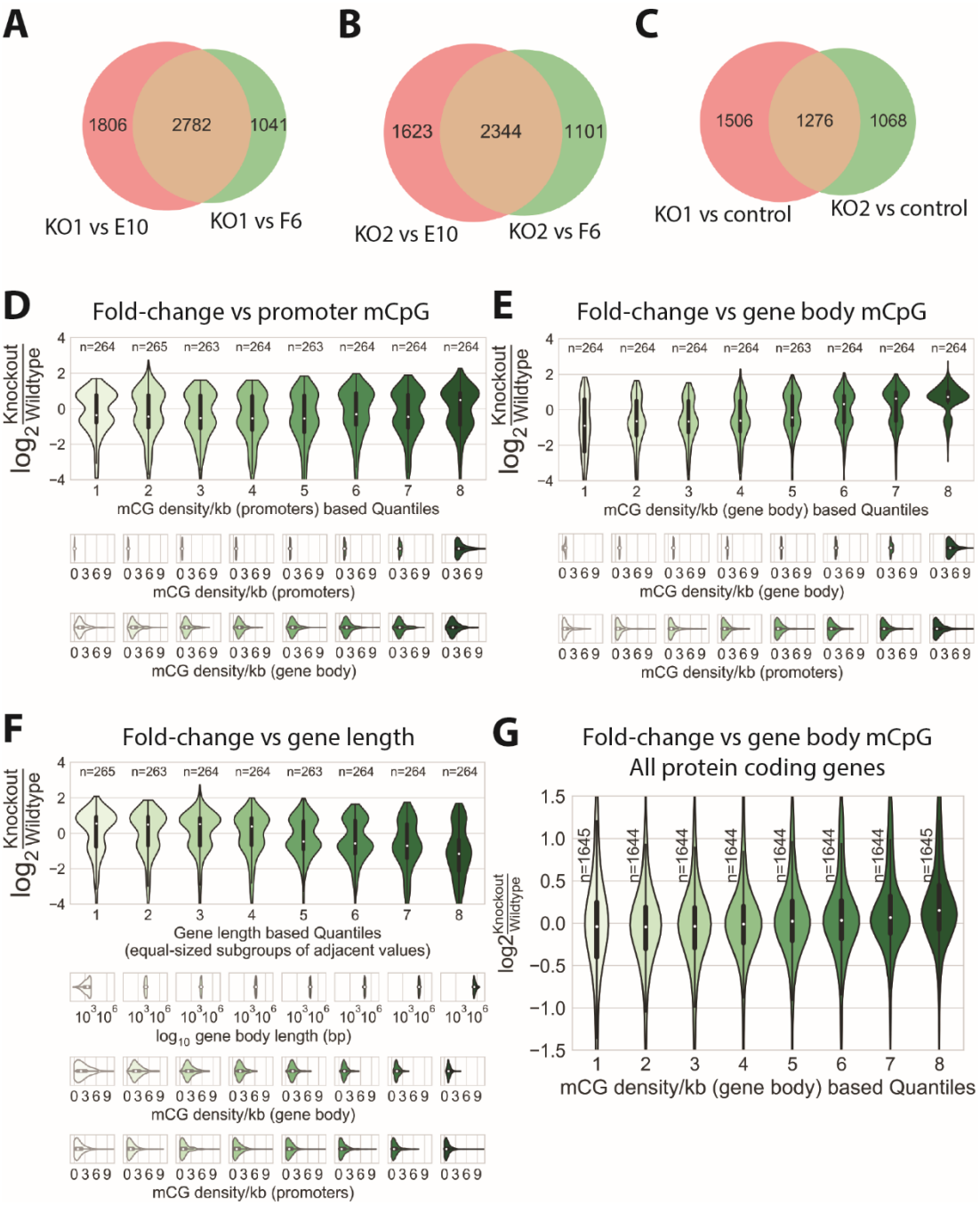


Figure 40. MeCP2 KO LUHMES-derived neurons exhibit gene body, methylation-dependent transcriptional changes.

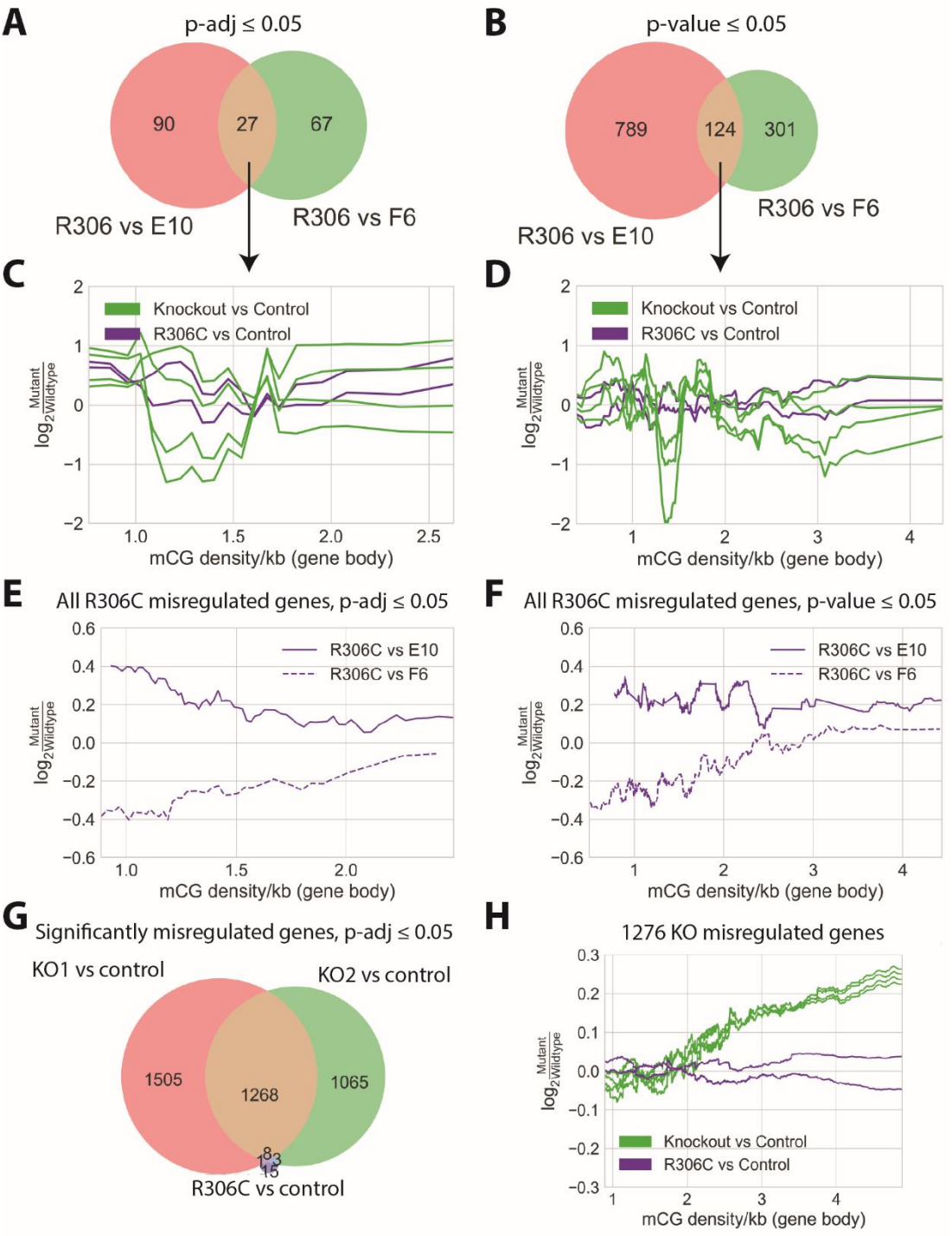
(A,B,C) Venn diagrams of WT and KO samples, using only significantly changing genes as determined with a p-adj cut-off of 0.05. **(A)** KO1 compared to 2 control cell lines. **(B)** KO2 compared to 2 control cell lines. **(C)** KO1 and KO2 compared to an averaged control sample (averaged data of E10 and F6). **(D,E,F)** Violin plots depicting gene expression changes of significantly changing genes between a collective KO and a collective WT sample, binned according to **(D)** promoter methylation, **(E)** gene body methylation or **(F)** gene length. Small plots below the main figure indicate the gene body mCG density, promoter mCG density and gene length ranges of genes within each violin plot. **(G)** Violin plots depicting gene expression changes of all protein coding genes (13154) between a collective KO and a collective WT sample, binned according to gene body methylation.

The precise mechanism that occurs downstream of MeCP2 binding to methylated DNA that contributes to gene repression is unknown. Recruitment of the HDAC-containing NCoR/SMRT corepressor complex throughout the gene body might alter the chromatin status of the gene in order to contribute to its repression. The deacetylation of histones by HDAC3 may then inhibit the progression of RNA polymerase (RNAP) through the gene. In addition, the occupation of methylated sites by direct binding of MeCP2 to mCG throughout the gene could itself hinder the progression of RNAP. Most likely, MeCP2 is exerting its effect on gene expression via a combination of these mechanisms. Patients with an R306C point mutation and *Mecp2*$^{R306C/y}$ mice have a less severe phenotype compared to other point mutations (Bebbington *et al.*, 2008; Brown *et al.*, 2016; Cuddapah *et al.*, 2014). MeCP2-R306C protein retains the ability to bind to sites of methylation within a gene body, but cannot recruit the NCoR/SMRT complex to these sites, perhaps resulting in an intermediate level of gene repression and a less severe Rett syndrome phenotype.

In order to test this hypothesis I performed RNA-sequencing analysis of the MeCP2-R306C LUHMES cell line in quadruplicate, again using E10 and F6 cell lines as controls. Analysis revealed, only 27 genes to be significantly misregulated in R306C neurons when compared to both control cell lines (Figure 41A). This measure of significance (p-adj, p-adjusted value) takes into account the variability between the individual replicates when calculating significance and was used for the KO RNA-seq data analysis. When the less stringent p-value measure of significance is used the number of significantly misregulated genes increases to 124 (Figure 41B). Assessment of the extent of gene misregulation for both sets of genes showed no gene-body mCG density-dependent effects (Figure 41C+D, purple lines). These same genes also showed no correlation in the KO datasets (Figure 41C+D, green lines).

Figure 41. MeCP2 R306C LUHMES-derived neurons show little transcriptional deregulation.

(A) Venn diagram depicting the significantly changing (p-adj ≤ 0.05) genes in R306C vs WT comparisons. **(B)** Venn diagram depicting the significantly changing (p-value ≤ 0.05) genes in R306C vs WT comparisons. **(C)** Line plot showing fold changes between KO and WT samples (green lines) and R306C and WT samples (purple lines) for the 27 significantly changing genes identified in A. **(D)** Line plot showing fold changes between KO and WT samples (green lines) and R306C and WT samples (purple lines) for the 124 significantly changing genes identified in B. **(E)** Line plot showing fold changes between R306C and WT samples for all significantly changing genes identified in A. **(F)** Line plot showing fold changes between R306C and WT samples for all significantly changing genes identified in B. **(G)** Venn diagram showing the overlap of gene misexpression between KO and R306C samples (Also see Supplementary Figure 13). **(H)** Line graph showing fold changes between KO and WT samples (green lines) and R306C and WT samples (purple lines) for the1276 significantly changing genes identified in the KO/WT comparison in Figure 40C.



As it is difficult to draw conclusions from such a small number of genes, the line plots were repeated to include all misregulated genes from each R306C vs control comparison (i.e the 117 genes in R306C vs E10 and the 94 genes in R306C vs F6 comparisons when the p-adj value of significance is used). This analysis failed to reveal any gene-body mCG dependent misregulation (Figure 41E), even when the larger p-value dataset was used that contains 425

and 914 genes (Figure 41F). In order to understand whether the small numbers of misregulated genes in the R306C dataset are also misregulated in the KO datasets, Venn diagrams comparing the significantly changing genes were drawn. This analysis revealed some overlap, but the majority of genes in the R306C dataset were not altered in the KO dataset (Figure 41G + Supplementary Figure 13). Finally, the 1276 significantly changing genes that were identified in the KO dataset (Figure 40C) were plotted using the R306C log_2FC values. Even though the vast majority of these genes are not significantly changing in the R306C dataset, perhaps they exhibit small shifts in expression which do not pass the test for statistical significance, but do demonstrate a mCpG-dependent effect. This analysis also revealed no correlation between gene expression change and gene body methylation (Figure 41H). Overall, it would appear that R306C LUHMES-derived neurons have extremely minor gene expression changes upon mutation of MeCP2, which do not correlate with the gene-body methylation density of the gene. The ability of R306C MeCP2 to still bind to sites of methylation throughout the genome suggests that perhaps MeCP2 exerts its repressive effect via physical binding and obstruction of RNA polymerase alone.

While a strong and consistent decrease in RNA level is observed in both MeCP2 KO and R306C LUHMES-derived neurons, gene expression defects are observed only in the MeCP2 KO cell lines. The extent of this gene misregulation appears to agree with a model whereby MeCP2 binds to methylated cytosines within the body of genes to contribute to their repression. Removal of MeCP2 from the gene body removes a block on the progression of RNA polymerase and thus causes an increase in the expression level of those genes. While the NCoR/SMRT complex is not implicated in this obstruction model of gene expression, the corepressor complex clearly has a role in the maintenance of RNA levels in neurons as R306C LUHMES-derived neurons and mouse brain regions have a loss of total RNA. Since the R306C point mutation causes Rett syndrome, this suggests that reduction of total RNA is a) independent of increased gene expression and b) extremely deleterious for brain function. It is unknown how the R306C point mutation can cause a decrease in total RNA but not have an effect on the transcriptome of protein coding genes. Perhaps, the loss of total RNA is predominantly due to a loss of rRNA, and the NCoR/SMRT complex can regulate the level of rRNA, but cannot affect the expression of MeCP2-bound protein-coding genes. Further work is needed to reveal the precise molecular mechanisms that cause these two transcriptional outputs.

6.5 Testing the model: overexpression of mutant MeCP2

6.5.1 Generation of LUHMES-derived neurons overexpressing mutant MeCP2

The R306C mutation causes little to no transcriptional defects when introduced into endogenous MeCP2 in LUHMES-derived neurons. This is in contrast to previous analysis of *Mecp2*$^{R306C/y}$ mouse cerebellum where small changes in gene misregulation were observed (Gabel *et al.*, 2015). In order to examine this in more detail, LUHMES cell lines were generated that overexpress mutant R306C MeCP2 on top of WT MeCP2. Based on the LUHMES R306C KI data, it would be expected that overexpression of R306C will result in a similar transcriptional profile to the overexpression of WT MeCP2. Alternatively, if R306C protein is truly non-functional its overexpression in LUHMES cells would result in a similar transcriptome to WT LUHMES cells.

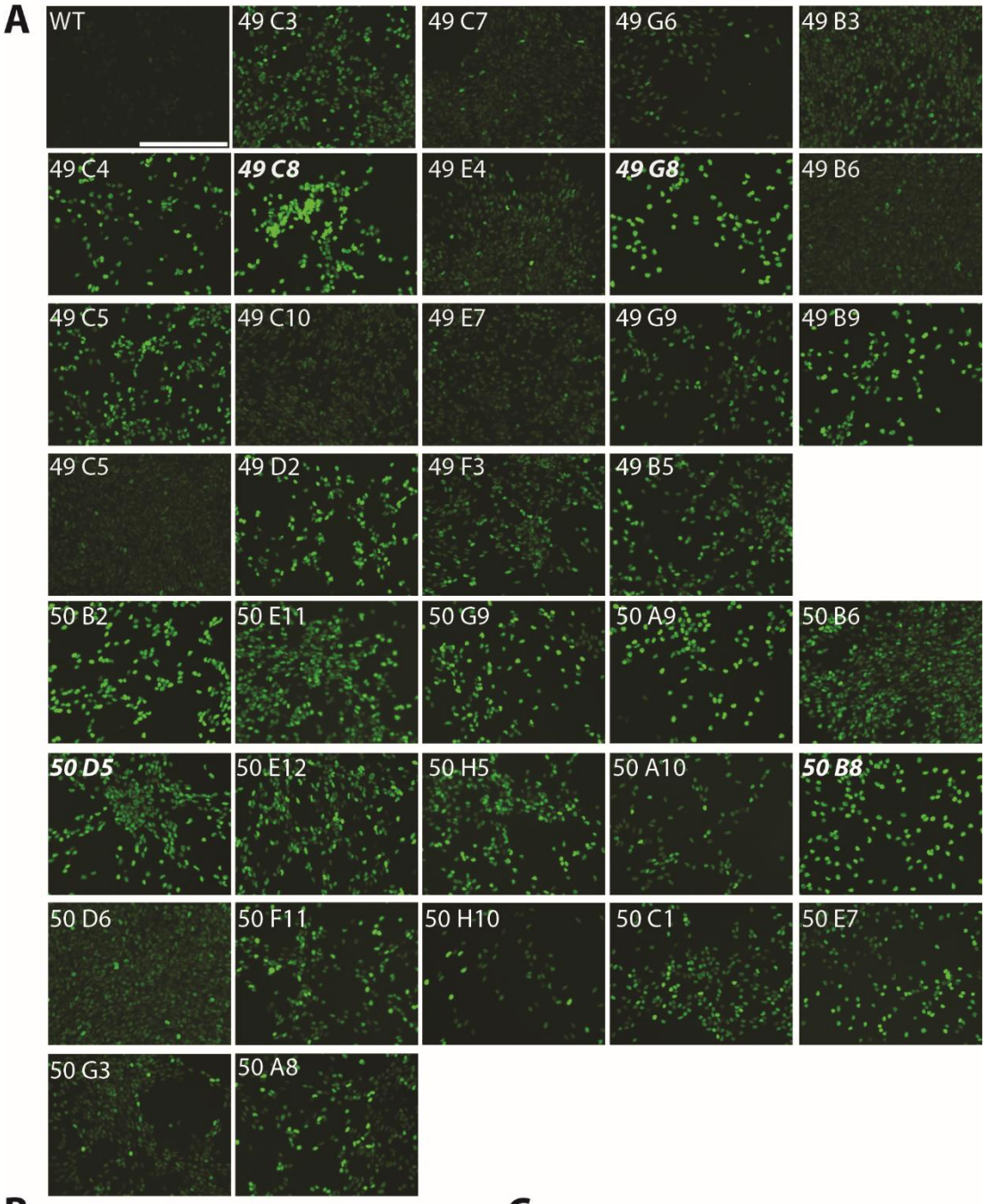
The positive control for this experiment is overexpression of WT MeCP2. The two negative controls are WT cells infected with a lentivirus containing no MeCP2 (OEC) and cells overexpressing of R111G, a protein that has reduced stability (this study, Figure 26), and reduced binding to methylated DNA (Heckman *et al.*, 2014)and thus should not cause any methylation-dependent gene misregulation compared to WT cells.

I made two batches of lentiviruses: virus 49 contained *R306C-MECP2* and virus 50 contained *R111G-MECP2*. After LUHMES cell infection and single-cell sorting, individual LUHMES clones were differentiated in 24-well plates, and probed for increased MeCP2 expression compared to a WT cell line. All clones had increased expression of MeCP2 compared to WT cells (Figure 42A), and two clones per virus infection were chosen for downstream analysis based on high and uniform MeCP2 expression. The mutant transgene was confirmed to be present in these four cell lines by sequencing analysis (Figure 42B).

Figure 42. Generation of mutant MeCP2 overexpression cell lines.

(A) Screening of cell lines for clones that have uniform, high levels of MeCP2 protein. Clones are differentiated in 24-well plates, immunostained with D4F3 MeCP2 primary antibody (Cell signalling) and imaged in the Incucyte machine. 49: cell lines infected with virus containing *MECP2-R306C*. 50: cell lines infected with virus containing *MECP2-R111G*. Cell lines in bold italic

writing were chosen for downstream analysis. **(B+C)** Sequencing of gDNA using cDNA specific primers.



6.5.2 Transcriptional defects in mutant overexpression cell lines

RNA-sequencing of four cell lines was performed in triplicate: OEC and OE CMV C10 as described in Supplementary Figure 10B, and R111G OE B8 and R306C OE C8 as described in Chapter 6.5.1. Western blotting of all samples that were sent for RNA-sequencing is shown in Figure 43A. Quantification of this shows that the overexpression of R111G-MeCP2 produces a total MeCP2 level of 6-fold WT level and is equivalent to the OE CMV C10 control (overexpression of WT MeCP2), while the overexpression of R306C reaches a total level of 9-fold (Figure 43B).

Using Venn diagrams to visualise the number of genes that are misregulated in both R111G OE and R306C OE cell lines reveals large numbers of genes that are misexpressed in all comparisons (Figure 43C). When comparing R111G OE to OEC, 3930 genes are misexpressed, suggesting that the presence of 6-fold R111G MeCP2 protein on top of WT protein affects gene expression (Figure 43Ci). This misregulation, however, does not mean that the overexpressed R111G protein is functioning as WT protein. Indeed, comparison of R111G OE to WT OE (which contain the same amount of total MeCP2 protein, Figure 43B) reveals that 5607 genes are misregulated (Figure 43Cii). There are fewer misregulated genes in the R306C OE cell line than the R111G OE cell line, but the pattern is the same in that neither mutant OE cell line has a transcriptome similar to WT cells or to WT OE cells. This indicates that both R111G and R306C, when overexpressed, retain some functionality to induce transcriptional changes, but both have lost complete WT function.

These large numbers of misregulated genes tell us that the overexpression of R111G and R306C affects the global transcriptional profile of neurons; but it does not tell us if this is mCpG-dependent and therefore whether the overexpressed protein is behaving in a manner that recapitulates WT MeCP2. In order to explore this, the log_2FC between the mutant OE cell lines and the OEC control cell line were plotted in relation to gene body mCpG density (Figure 43D). This analysis revealed a minor relationship between gene-body mCpG density and log_2FC for R111G OE, indicating that although 3930 genes are misexpressed, these are not misexpressed in a gene-body mCpG-dependent manner. The overexpression of R306C, on the other hand, has a relationship whereby the more gene-body mCpG density a gene has, the more repressed it is in the overexpression cell line. This suggests that the R306C protein, despite the mutation, retains some capabilities to act like WT protein and bring about gene repression in a methylation-dependent manner. This repression, however, is not as pronounced as when WT MeCP2 is overexpressed; which leads to 5448 misregulated genes

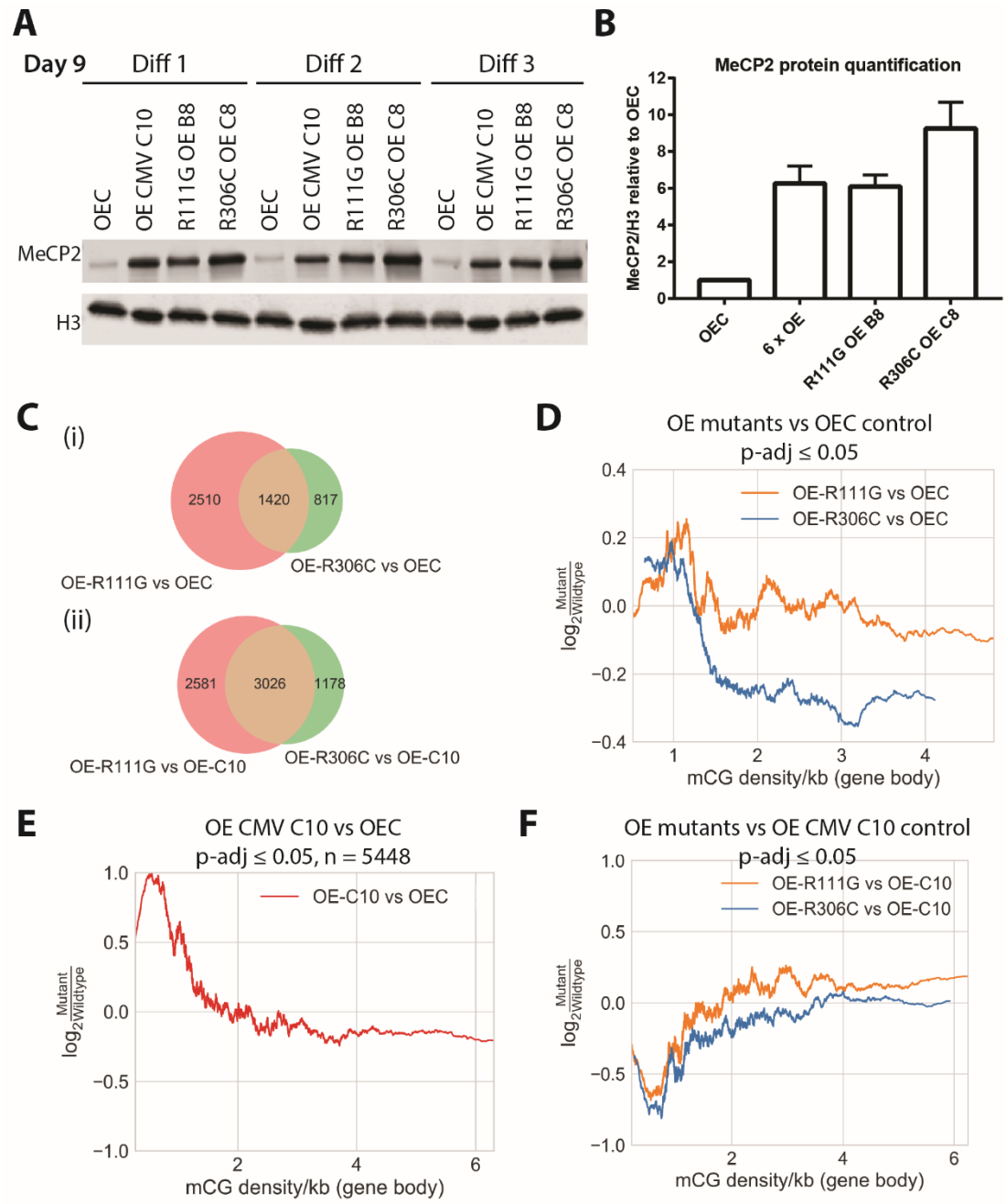


Figure 43. Analysis of mutant overexpression cell lines.

(A) Replicate differentiations of cell lines overexpressing R111G and R306C along with control cell lines, probed with D4F3 (Cell Signalling) antibody. **(B)** Quantification of blot in C. H3 is shown as a loading control for all Western blots. **(C)** Venn diagram comparing significantly misregulated genes (p-adj ≤ 0.05) in R111G OE and R306C OE when compared to (i) OEC or (ii) OE CMV C10. **(D)** Line plot of gene expression changes relative to gene body mCpG density in R111G OE and R306C OE lines compared to OEC. **(E)** Line plot of gene expression changes relative to gene body mCpG density in OE CMV C10 cell line compared to OEC. **(F)** Line plot of gene expression changes relative to gene body mCpG density in R111G OE and R306C OE lines compared to OEC CMV C10.

(as opposed to 2237 for R306C OE) and displays a strong relationship with gene-body mCpG density (Figure 43D: compare y-axis scales in Figure 43D+E). Thus, despite being expressed to a higher level, R306C cannot repress genes to the same degree as WT protein. This can be shown by plotting the gene-body mCpG densities against $\log_2$ fold-changes between WT OE and R306C OE cell lines (Figure 43F). This graph demonstrates that the higher the gene-body mCpG density of a gene, the more expressed it will be in the R306C OE sample compared to the WT OE sample, i.e. genes with higher gene-body mCpG densities are more strongly repressed when WT protein is overexpressed than when R306C protein is overexpressed. The same is true for the overexpression of R111G protein (Figure 43F). This suggests that neither mutant protein can function as WT and that both are impaired in their ability to repress gene expression in a gene-body mCpG dependent fashion.

Thus the data supports a model where R306C protein is non-functional, and its overexpression in LUHMES-derived neurons does not induce significant methylation-dependent gene expression changes. This is in line with published data demonstrating transcriptional defects in mice that express R306C as an endogenous locus KI (Gabel *et al.*, 2015), and the lack of a phenotype in mice that overexpress R306C protein (Heckman *et al.*, 2014). The data here, however, demonstrates that neither R111G or R306C are completely inert, non-functional proteins as their overexpression does cause large numbers of genes to be misexpressed compared to WT cells. Further analysis of these RNA-sequencing datasets will be required to determine if there are any common gene properties to these misexpressed genes.

6.6 Discussion

This chapter analysed the morphological and transcriptional phenotypes in MeCP2 KO and R306C LUMHES-derived neurons. The finding that LUHMES neuronal progenitors can differentiate into neurons as determined by immunocytochemistry and extensive RT-qPCR analysis confirms, in combination with published datasets, that RTT is not a neurodevelopmental disorder. Despite observations that neurons differentiated from RTT patient iPSCs display reduced TuJ staining (Kim *et al.*, 2011), and that olfactory epithelium neurons have an apparent neuronal maturation defect (Gabriele V. Ronnett *et al.*, 2003), girls who are born with RTT mutations can survive well into adulthood and despite significant neuronal malfunction, mature neurons do develop (Dawna Armstrong, 2005). A reduction in TuJ staining has not been observed in other systems (Ananiev *et al.*, 2011; Caballero *et al.*, 2009; Yazdani *et al.*, 2012), and the maturation defect in olfactory epithelial neurons was

observed in samples from RTT patients who had no detectable MeCP2 mutation, suggesting that this effect is a pathological outcome in patients, but not necessarily a primary consequence of MeCP2 mutation. This study found no decrease in the expression level of TuJ nor of MAP2, a microtubule-associated protein. This is in contrast to published reports of decreased mRNA (Colantuoni *et al.*, 2001) and protein levels (Kaufmann *et al.*, 2000) of this microtubule-stabilising protein in post-mortem brain samples from RTT patients. Mining of the LUHMES KO vs control RNA-sequencing dataset reveals a small but statistically significant reduction in *MAP2* mRNA levels in one of the two KO cell lines (KO1 vs E10: log_2FC 0.12, p-value 0.26, KO1 vs F6: log_2FC 0.052, p-value 0.62, KO2 vs E10: log_2FC -0.25, p-value 0.016, KO2 vs F6: log_2FC -0.32, p-value 0.002). Nonetheless other tissue-culture models report normal levels of MAP2 (Li *et al.*, 2013c; Marchetto *et al.*, 2010), thus reduced MAP2 levels are not a consistent defect observed in RTT models. Historically RTT has been considered a neurodevelopmental disorder, but the observation that symptoms can be reversed in a mouse model (Guy *et al.*, 2007) and recent evidence from a variety of model systems that mature neurons can be differentiated from precursors suggests that RTT is not a disorder of neuronal development. The analysis performed in LUHMES cells in this study further confirms the lack of a differentiation defect and establishes normal developmental dynamics by analysis of 22 markers by RT-qPCR.

While MeCP2 mutant cells are reported to differentiate into mature neurons, the neurons seem to exhibit a vast array of morphological defects including reduced nuclear size and decreased neurite length (Supplementary Table 3 and
Supplementary Table 4). Smaller neurons, shorter neurites and reduced dendritic complexity have also been observed in post-mortem brain samples from patients with Rett syndrome (Dawna Armstrong, 2005; Dawna Armstrong *et al.*, 1995). The finding here that MeCP2 KO neurons do not have shorter neurons or smaller nuclei could reflect the young age of neurons that were assessed. Analysis of older neurons may reveal a deficit in neurite length; however the technique should be altered so that the Incucyte software is able to detect individual neurites within the thick mesh of neuronal projections. To do this, LUHMES cells could be infected at a very low multiplicity of infection (MOI) with a lentivirus expressing GFP, such that only a few neurons in a single field of view are GFP-positive (Scholz *et al.*, 2011). This would enable the Incucyte software to detect individual neurons and calculate the neurite length of GFP-positive neurons in an accurate manner. Similarly, analysis of the nuclear volume of older LUHMES-derived neurons using a more sophisticated approach such as ChromATin (Linhoff *et al.*, 2015) may reveal nuclear volume decreases.

The finding here, that gross morphological changes are not apparent, whereas strong and robust transcriptional changes can be detected suggests that the morphological changes that are frequently observed in mutant neurons are a downstream effect of global transcriptional changes. Assessment of young neurons (only 9 days after differentiation was initiated) has enabled the primary transcriptional consequences of MeCP2 mutation in human neurons to be analysed. RNA-sequencing analysis at this stage has not revealed the presence of a small number of MeCP2 target genes, but rather that large numbers of genes are mis-expressed upon MeCP2 deletion. While it may seem obvious to state that transcriptional changes when we delete a methylated DNA-binding protein are upstream of morphological changes, this data has implications for therapeutic strategies for RTT. While nonselective drugs such as ketamine that affect a wide variety of neurological pathways are being trialled for amelioration of RTT symptoms (Katz *et al.*, 2016), the fact that many tested drugs so far have fallen below the expected efficacy standard (for example mecasermin/IGF-1 (Khwaja *et al.*, 2014)) suggests that the broad transcriptional changes (and resulting neurological phenotypes) are too varied for a single drug that targets an individual pathway to provide significant levels of benefit (Kaufmann *et al.*, 2016).

The analysis here found that gene body methylation, rather than promoter methylation, is the primary modification through which MeCP2 exerts its repressive effect on transcription, in agreement with previous reports in mice (Kinde *et al.*, 2016). While the role of promoter methylation is well established, the role of gene body methylation is not so clear (Jones, 2012; Moore *et al.*, 2013). Gene body methylation has been associated with H3K36me3 and actively transcribed genes (Baubec *et al.*, 2015), suppression of alternative tissue-specific intragenic promoters (Maunakea *et al.*, 2010), and with gene repression during neuronal development (Lister *et al.*, 2013). Precisely how MeCP2 is exerting its effect on transcription via binding throughout gene bodies is unknown, but may occur via modulation of the initiation and/or elongation of RNA polymerase (RNAP).

There have been few analyses of the dynamics of RNAP progression in MeCP2 mutant samples. A recent study found the total level of Pol II Serine5-phosphorylation to be decreased *Mecp2-null* neurons in brain slices of *MeCP2*$^{+/-}$ heterozygous female mice, which could indicate decreased rates of Pol II initiation and therefore explain the global decrease in RNA amount observed in this study (Linhoff *et al.*, 2015). However the total amount of Pol II in the WT and mutant samples were not measured using the same sensitive technique, and thus it cannot be ruled out that total RNAP levels are decreased in mutant samples and

therefore all RNAP modifications are reduced (Linhoff *et al.*, 2015). An alternative study found increased Pol II binding by ChIP-qPCR at promoters of immediate-early genes that display increased in expression in *Mecp2*^{T308} mice (Su *et al.*, 2012). This study saw an increased level of Ser2-phos and Ser5-phos to the same degree thus suggesting, at least at these two immediate-early response genes, MeCP2 modulates the ability of Pol II to bind to promoters, not just the elongation rate of bound Pol II. Analysis of Pol II modification status in LUHMES cells combined with histone modification analysis may elucidate the mechanism through which MeCP2 affects transcription. In particular, H3K36me3 and RNAP II Ser2-phos levels would assess the efficiency of polymerase progression through the gene body, while RNAP II Thr4-phos levels will determine if transcripts are efficiently terminated in the absence of MeCP2 (Schlackow *et al.*, 2017). Combining Pol II and histone modification levels with studies assessing the rate of nascent transcription will be critical to deciphering the mechanism of gene-body, MeCP2-mediated transcriptional repression.

In addition to such endogenous studies of the transcriptional and chromatin landscape in LUHMES-derived neurons, more artificial studies could help to elucidate the importance of gene-body MeCP2 binding on transcription. For example, MeCP2 could be tethered to sites throughout the body of endogenous genes in WT cells (perhaps by Cas9-recruitment), and the transcriptional output of this locus measured by RT-qPCR analysis. By performing this experiment on loci that have different endogenous transcriptional outputs (e.g. highly expressed vs lowly expressed genes), different endogenous methylation profiles, and that are located within alternative chromatin environments, the ability of MeCP2 to repress transcription and how the local chromatin environment affects this repression can be assessed in the endogenous environment of human neuronal cells. In addition, the combination of MeCP2 tethering assays with HDAC inhibitor application (such as trichostatin A) would determine the importance of histone deacetylation activity for the repression of endogenous genes in their endogenous chromatin environments. This analysis could be performed in LUHMES-derived neurons or even in LUHMES neuronal precursors as MeCP2 is expressed and total RNA defects have been observed at this early stage. If such experiments were successful, the contribution of RTT mutations (such as R306C) to this repression could then be assessed.

The finding that in LUHMES cells, deletion of MeCP2, insertion of the R306C point mutation, or overexpression of MeCP2 by 12-fold results in a decrease of RNA in proliferating neuronal precursors is somewhat surprising. Analysis of total RNA levels

during hESC differentiation found no change in RNA amount at the neural precursor stage, and only saw a decrease in neurons that were 2-weeks and 4-weeks old (Li *et al.*, 2013c). Most transcriptional studies are performed on mature neurons, either from mouse brain or tissue culture differentiation systems, but the finding that RNA levels are affected even at the precursor stage suggests that transcriptional studies should be assessed at this early stage as well. If transcriptional misregulation is not observed in LUHMES precursors then the decrease in RNA amount cannot be explained by secondary effects of MeCP2 loss of function and experiments to determine the mechanism leading to the loss of RNA will be vital.

The observation that total RNA is reduced in both MeCP2 KO and R306C LUMHES-derived neurons is somewhat contradictory for a protein that is hypothesised to repress transcription. Nevertheless this observation agrees with data from MeCP2-null human (Li *et al.*, 2013c) and mouse (Yazdani *et al.*, 2012) ESC-derived neurons as well as *in vivo* in olfactory neuroepithelium (Rube *et al.*, 2016), suggesting it may play an important role in RTT pathology. The majority of RNA is composed of ribosomal RNA (rRNA), and thus perhaps the discordance between the apparent repressive effect of MeCP2 by RNA-seq analysis and the apparent activation effect as judged by HPLC could explained by effects on different RNA populations. In order to assess the contribution of rRNA and mRNA to the decrease in total RNA levels, rRNA and pre-rRNA levels could be assessed by Northern blot or qPCR analysis. Levels of the 5.8S and 5S rRNAs were found to be decreased in neurons differentiated from hESCs indicating a strong reduction in rRNA, but not ruling out a similar reduction in mRNA (Li *et al.*, 2013c). In line with this data, significantly smaller nucleoli (the sites of rRNA transcription) have been observed in tissue cores from *MeCP2*$^{-/y}$ brain and in cultured cortical neurons suggesting reduced transcription of rRNA (Singleton *et al.*, 2011). Analysis of the number and size of nucleoli in LUHMES-derived mutant neurons, for example by staining for Ki67, nucleolin or using the CytoPainter Nucleolar/Nuclear staining kit from Abcam, may contribute towards a greater understanding.

A potential hypothesis for how MeCP2 loss results in decreased RNA levels could involve MeCP2 playing a role in the repression of spurious transcription throughout gene bodies. MeCP2 bound to gene body methylation could inhibit RNA polymerase initiation at these sites, but allow the elongation of existing polymerase molecules through the gene, a proposed mechanism for the purpose for gene body methylation (Jones, 2012). The removal of MeCP2 would result in increased spurious transcription throughout the gene body, which

could result in inefficient transcription of the full-length transcript and thus decreased levels of the intended transcript (

Figure 44). Experiments to test this hypothesis would involve analysis of RNAP II by ChIP-seq and assessments of its modification status by mNET-seq, in particular paying attention to the Ser5-phos, Ser2-phos and Thr4-phos modifications which mark the transcription initiation, elongation and termination sites of genes (Heidemann *et al.*, 2013; Nojima *et al.*, 2015; Schlackow *et al.*, 2017). Techniques to analyse nascent RNA transcription, such as 4sU-seq (Cleary *et al.*, 2005), Gro-seq (Core *et al.*, 2008) and TT-seq (Schwalb *et al.*, 2016) will be useful for analysing RNAP progression through genes and directly assessing sites of spurious transcription. This could be combined with inhibition of the exosome complex to increase the chance of observing spurious transcription and transient RNA molecules. Alternatively, the total RNA decrease could simply be a secondary effect, caused by more fundamental transcriptional changes that lead to smaller cells and subsequent decreased transcriptional outputs (Padovan-Merhar *et al.*, 2015). While it is incongruous to associate a transcriptional repressor protein with maintenance of high RNA levels, further experiments to determine the mechanistic cause of this phenotype will be necessary to determine primary versus secondary effects.

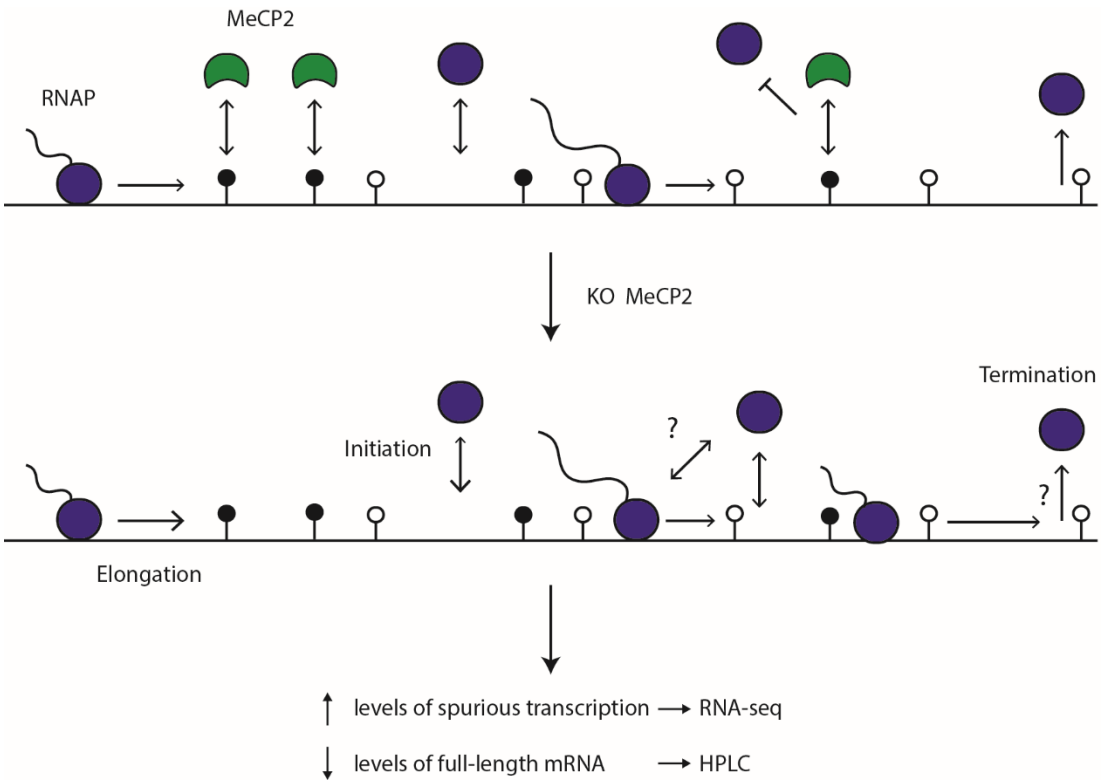


Figure 44. Hypothetical mechanism for MeCP2-mediated gene regulation.

This study found that 9-fold overexpression of R306C-MeCP2 resulted in 4204 genes that were misexpressed when compared to the overexpression of WT MeCP2, suggesting that R306C-MeCP2 does not function like the WT protein. Closer analysis of these genes revealed that genes with higher mCpG gene-body densities were more highly expressed in R306C OE than WT OE, suggesting impairment in the ability of MeCP2 R306C to repress gene expression in a gene-body methylation dependent manner. On-the-other-hand, comparison of R306C OE to WT cells, revealed a weak ability of R306C MeCP2 to repress genes in a gene-body methylation dependent manner. Overexpression of R306C MeCP2 to 9 times the level of the WT protein results in this extra protein binding to and occupying methylated sites in the genome that are not already covered by WT MeCP2. MeCP2 binding is not a static event, and the residency time of WT MeCP2 is approximately 15 seconds as measured by fluorescence recovery after photobleaching (FRAP) (Schmiedeberg *et al.*, 2009). Thus, the law of mass action applies, where extremely high levels of MeCP2 protein in the nucleus forces greater and more frequent binding to DNA. This binding is enough to bring about gene repression in a methylation-dependent manner, but the inability of R306C-MeCP2 to recruit NCoR/SMRT may reflect the inability of overexpressed R306C-MeCP2 to repress transcription to the same degree as overexpressed WT MeCP2. The recent observations that R306C protein binds to methylated DNA with a reduced affinity, despite retaining a completely intact MBD (Brown *et al.*, 2016; Heckman *et al.*, 2014) would suggest that R306C in LUHMES cells not only affects transcription by a lack of recruitment of NCoR/SMRT, but also by reduced binding to DNA. Indeed, recent ChIP-qPCR data from Dr Justyna Cholewa-Waclaw has demonstrated that R306C has an approximately 50% reduction in DNA binding in LUHMES-derived neurons (personal communication). This data suggests that the LUHMES R306C KI RNA-sequencing data here should be more pronounced. Perhaps performing the experiment using more R306C clones will reveal a more pronounced defect, or perhaps analysis of nascent transcription using the techniques described above will be more fruitful.

To conclude, MeCP2 null and R306C LUHMES-derived neurons have been phenotypically assessed by a number of parameters. Thus far, no defect in neuronal morphology has been observed. Re-analysis of neuronal volume and neurite length using more sophisticated techniques might reveal morphological abnormalities. Analysis of neuronal function using the calcium imaging and electrophysiological techniques described in Chapter 3.6 will be useful to assess the ability of LUHMES-derived neurons to function in the presence of gross transcriptional defects. RNA-sequencing of MeCP2 KO neurons has revealed large numbers

of misregulated genes that are modulated in a gene-body methylation-dependent manner. Analysis of MeCP2 knock-down and overexpression LUHMES cell lines show similar and reciprocal gene expression effects respectively thus highlighting the functional significance of this mode of action for MeCP2 (J Cholewa-Waclaw + RR Shah *et al.*, unpublished data). Future experiments to assess the interplay of MeCP2, nascent transcription and RNA polymerase will be critical to understanding the precise molecular mechanism that underlies this repression. The observation that total RNA is dramatically reduced in neuronal precursors and neurons from MeCP2 null and R306C LUHMES cell lines is exciting, but also surprising. Further experiments need to be performed to assess if this is a consequence of sub-standard, smaller neurons, or if in combination with the RNA-sequencing defects, it contributes to RTT pathologies.

Chapter 7

7. Towards a protein transduction therapy for Rett syndrome

7.1 Introduction

For therapeutic intervention in disease, the direct application of protein or nucleic acid molecules to cells or tissues is generally not possible due to the impermeable nature of the plasma membrane. Viruses, however, such as AAV or lentiviruses readily infect a wide variety of cell types and can be manipulated to package a gene of interest such that application of the recombinant virus to cells will result in delivery and expression of that gene as required. The field of gene therapy has focussed primarily on rAAV vectors and has demonstrated delivery and long-term transgene expression, in a variety of cell types *in vivo* including muscle, liver, the eye, lungs and the brain (Samulski & Muzyczka, 2014). There are, however, some drawbacks to the *in vivo* use of AAV. The small packaging size restricts the use of gene therapy to small genes and limits the ability to control gene expression through the use of endogenous promoters and regulatory elements. Furthermore, a number of clinical trials have observed immune responses to the AAV vector or the transgene itself, in some cases resulting in clearance of the virus from the system (Halbert *et al.*, 1998; Manno *et al.*, 2006; Murphy *et al.*, 2008; Vandenberghe *et al.*, 2006). This affects the ability of rAAV to be delivered in multiple doses for long-term transgene expression and life-long therapeutic intervention, although is perhaps not so problematic for immune-privileged tissues such as the brain and eye (Bennett *et al.*, 2012; Daya & Berns, 2008).

The delivery of protein into cells (protein transduction) as an alternative to virus-mediated gene delivery is attractive for multiple reasons. Firstly, the dosage of protein can be more tightly controlled. Second, the insertional mutagenesis risk that viral vectors pose is not a

problem. Third, protein persistence has been observed for at least 2 months after topical application to mice (Wang *et al.*, 2013b). Fourth, initial studies in patients have demonstrated a lack of immune clearance of the protein in the blood, thus allowing for multiple dose regimes to be implemented without a loss of efficacy (J. Edmond Wraith *et al.*, 2008). But there have been fewer clinical trials using protein therapy compared to rAAV, thus more evidence is needed to rigorously test the latter point.

There are two alternative approaches to protein transduction; mixing the protein of interest with a substance that aids transduction, or tagging it with a peptide that can mediate plasma membrane translocation. Examples of the former include 10% carboxymethylcellulose (Wang *et al.*, 2013b), NaCl (Jr. *et al.*, 1978), sucrose and PEG (Okada & Rechsteiner, 1982), chloroquine (Luthman & Magnusson, 1983), and NaCl with propanebetain (D'Astolfo *et al.*, 2015). The observation that some proteins have natural cell-penetrating abilities, conferred by small regions within the protein paved the way for cell penetrating peptides (CPPs) to be identified, for example the homeodomain from the *Drosophila* Antennapedia protein (Derossi *et al.*, 1994), and a portion of the TAT protein from HIV-1 (Vives *et al.*, 1997). Fusion of these CPPs to heterologous proteins, such as cre recombinase and β-galactosidase, allows the translocation of the protein into cells and tissues and has been demonstrated to work *in vivo* (Kim *et al.*, 2009; Peitz *et al.*, 2002; Schwarze *et al.*, 1999).

While the sequence of CPPs varies, a common feature between all CPPs is that they are basic proteins. Although the exact mechanisms for transduction are unknown, the observation that polycationic peptides can also function as CPPs suggests that proteins that are naturally rich in lysine and arginine residues might have intrinsic protein transduction ability (Kabouridis, 2003). In line with this, cre protein on its own is able to transduce cultured cells and this activity is enhanced if cre is tagged with 6xhistidine or NLS signals (Lin *et al.*, 2004; Will *et al.*, 2002). MeCP2 is a very basic protein and has an intrinsic histidine run (Nan *et al.*, 1996), thus the hypothesis is that MeCP2 itself will be able to transduce cells without the need for a CPP.

This chapter aimed to determine if MeCP2 is able to cross the plasma membrane of MeCP2-null LUHMES-derived neurons. Such studies would lay the groundwork for future work towards a protein transduction therapy for Rett syndrome.

7.2 MeCP2 protein itself can cross the cell plasma membrane

Previous experiments performed by Dr Matthew Lyst have suggested that WT human MeCP2_e1 protein prepared from *E. coli* is able to transduce mouse embryonic fibroblasts (MEFs) in the presence of chloroquine. In his experiments, 4 hours after MeCP2 and chloroquine were added to cell culture media, Western blotting detected the presence of MeCP2 in nuclear extracts prepared from MEF cells, which do not express MeCP2. Furthermore, MeCP2 protein was only detected if the nuclear extracts were treated with benzonase, indicating that MeCP2 was only released into solution upon nucleic acid digestion, implying that it was bound to DNA.

Using MeCP2 prepared by Dr Matthew Lyst (Figure 45A) I added protein with or without chloroquine to 6-day old, MeCP2 KO, LUHMES-derived neurons. Nuclear extracts were prepared using benzonase, and probed for the presence of MeCP2 by Western blot analysis. Figure 45B shows that, similar to MEFs, MeCP2 protein was able to enter the nuclei of LUHMES-derived neurons within 4 hours. Yet, in contrast to MEFs, this transduction was not dependent on chloroquine. Figure 45C shows that MeCP2 protein was still present in nuclei 24 hours after initial protein addition to cell media. While this experiment suggests that MeCP2 protein is able to cross the cellular membrane and enter nuclei, the protein could be entering during the cell lysis and nuclei preparation steps. In order to assess this possibility, I performed immunofluorescence analysis using LUHMES-derived neurons. Figure 46A shows nuclear localisation of MeCP2 protein in WT neurons, and the absence of any signal in the KO or KO + chloroquine samples. While there is substantial MeCP2 immunoreactivity in the KO samples that were incubated with MeCP2 protein, the signal is not exclusively nuclear localised and a large majority of it appears to be extra-cellular. Further assessment of these images demonstrates that localised circles of MeCP2 protein are co-localising with foci of small, condensed or fragmented DAPI (Figure 46B, open arrows). These DAPI foci are often devoid of any MAP2 or NF staining, suggesting that these neurons are dead. On the other hand, intact DAPI foci that retain MAP2 and NF staining have very little co-localised MeCP2 (Figure 46Bii, box). In fact these neurons appear to have foci of MeCP2 protein that are sequestering on the outer surface of the cell body (Figure 46Bi, closed arrow), perhaps reflecting MeCP2 protein that is transiting into the cell, or an inability of protein to enter the neurons. Similar foci can be observed along the length of multiple neurites (Figure 46Bi, rectangles).

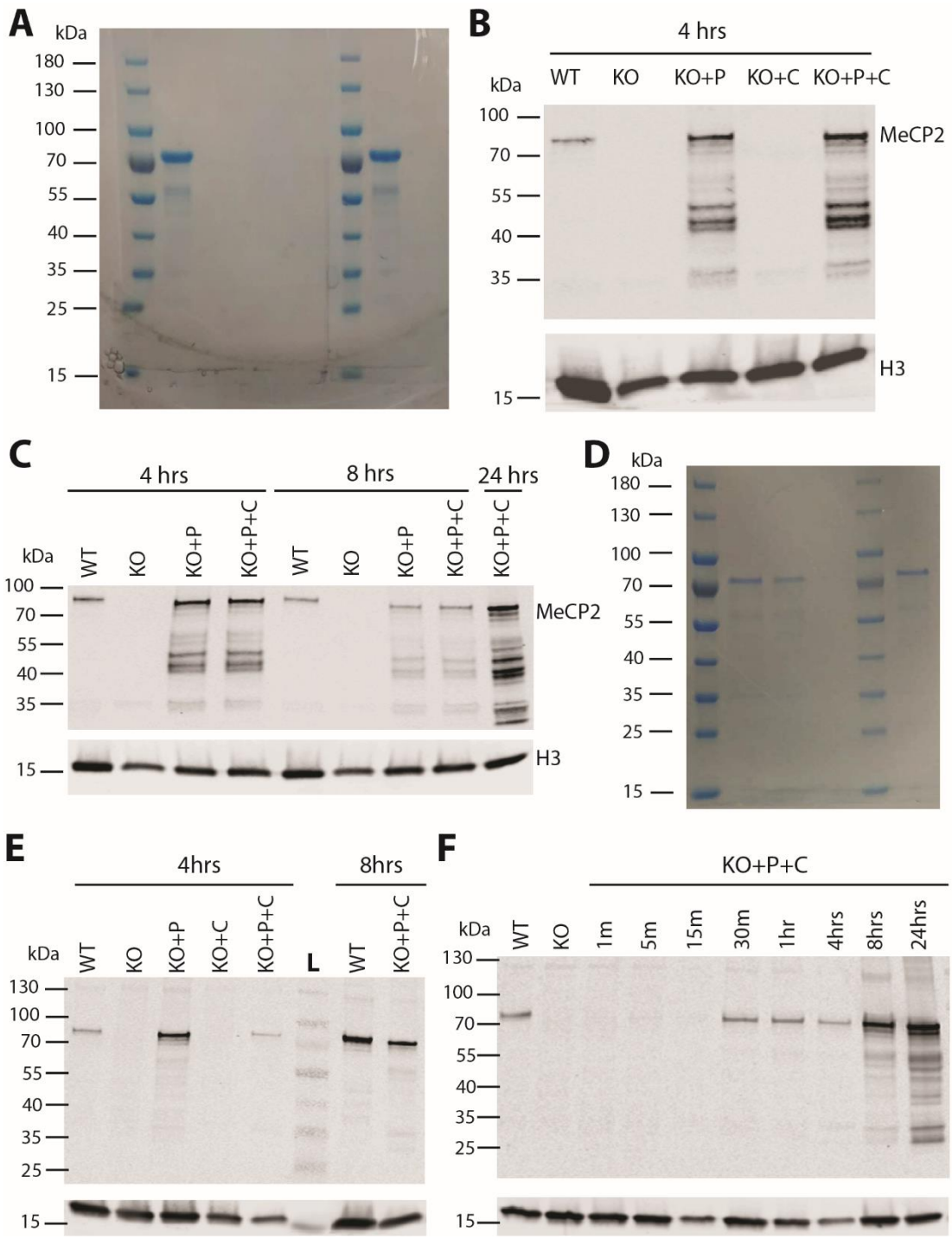


Figure 45. Analysis of protein transduction in LUHMES-derived neurons by Western blot.

(A) Coomasie staining of the 1st MeCP2 protein preparation by Dr Matthew Lyst. **(B)** Western blot analysis of protein transduction after 4 hours using the protein preperation shown in (A). **(C)** Timecourse of protein transduction using the protein preperation shown in (A). **(D)** Coomasie staining of the 2nd protein preparation by Dr Matthew Lyst. **(E)** Western blot analysis of protein transduction after 4 hours and 8 hours using the protein preparation shown in (D). **(F)** Timecourse experiment using the protein preparation shown in (D). H3 is probed as a loading control for all Western blots. MeCP2 antibody is M7443 (Sigma). P = MeCP2 protein. C = chloroquine. L = ladder.

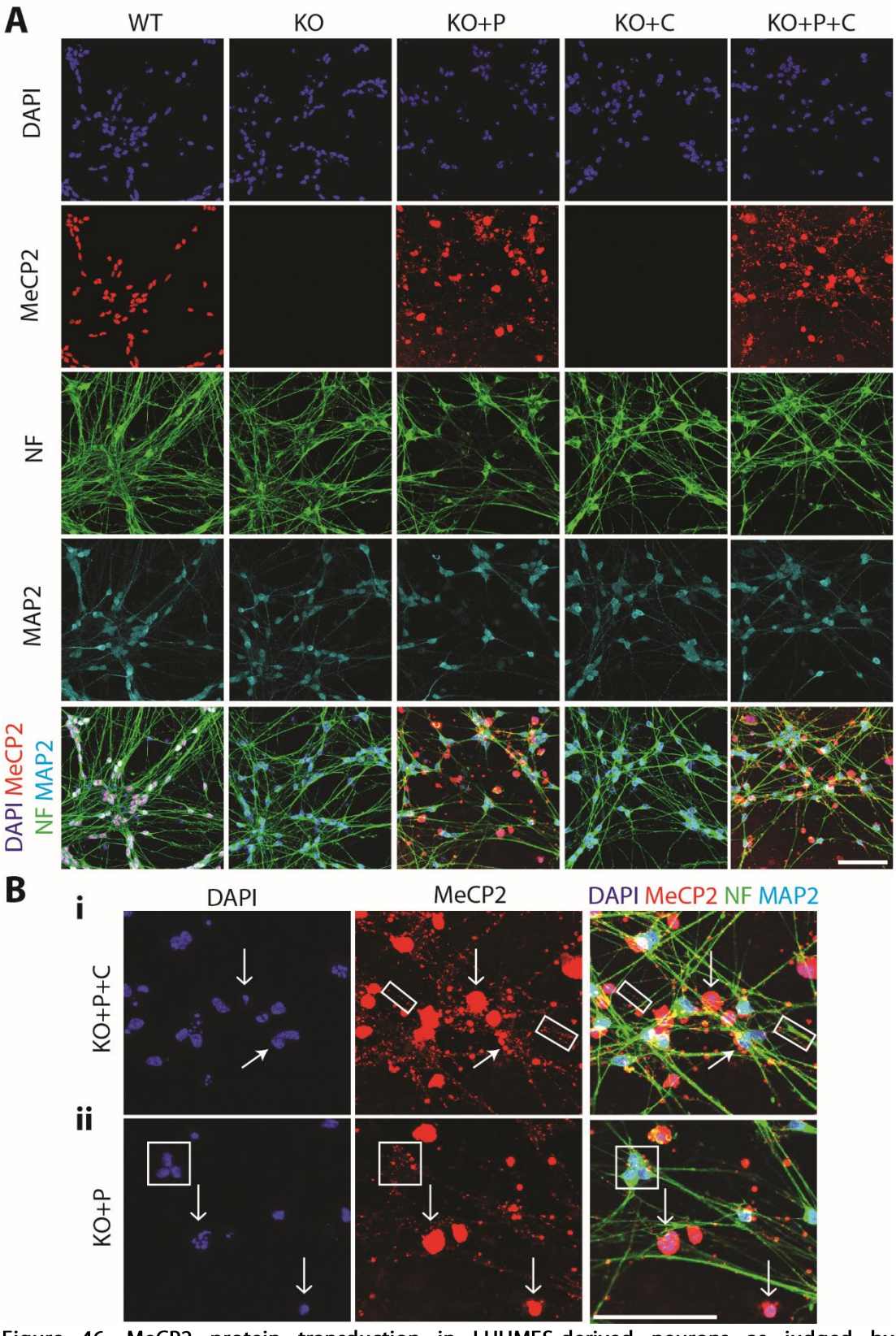


Figure 46. MeCP2 protein transduction in LUHMES-derived neurons as judged by immunofluorescence.

(A) IF images of LUHMES-derived neurons taken on day 9, after 4 hours of exposure to either MeCP2 protein (P), chloroquine (C) or both (P+C). MeCP2 antibody is D4F3 (Cell Signalling). Scale

bar is 100 µm. **(B)** Magnified images of MeCP2 KO neurons that were incubated with MeCP2 protein, either in the presence (i) or absence (ii) of chloroquine. Scale bar is 100 µm.

This data suggests that neurons which are transduced by MeCP2 protein are subsequently dying. Alternatively, it could be that only dying neurons are capable of being transduced. While intact nuclei were harvested for Western blot analysis, and full-length MeCP2 was observed in these samples, the ladder of MeCP2 immunoreactivity suggests that these neurons were successfully transduced by MeCP2 and then subsequently enter an apoptotic pathway (Figure 45B+C). In this experiment, there could be two stimuli for an apoptotic pathway. Perhaps there are contaminants in the MeCP2 protein preparation that affects the cells. Alternatively, the large quantities of protein that are being added to the neurons could be a trigger. In this experiment, 12.5pg of MeCP2 was added per neuron (Table 8, Chapter 2.26), which is approximately 10-fold higher than the amount of MeCP2 protein that exists in the nuclei of NeuN+ nuclei from mouse brain (Skene *et al.*, 2010) and 50-fold higher than the amount calculated for LUHMES-derived neurons (Table 9, Chapter 3.4.1).

In an attempt to avoid excessive amounts of cell death as observed by immunofluorescence, a new protein preparation was made by Dr Matthew Lyst. The first protein prep (Figure 45A) was purified with a Ni-NTA column using the endogenous Histidine run in MeCP2 (amino acids 366-372) while the second protein prep (Figure 45D) was made using shallower gradients of ionic strength/imidazole in order to better separate MeCP2 from contaminants. As can be seen in Figure 45E, 1.25pg/cell of this new protein was able to transduce LUMHES-derived neurons within 4 hours, and in the absence of chloroquine. Furthermore, the extra step in the protein purification strategy produced a purer preparation of MeCP2 (Figure 45, compare A to D), and this resulted in less cellular degradation products after protein transduction (Figure 45, compare C to E). A time course analysis established that MeCP2 could be detected in nuclear extracts 30 minutes after the addition of protein to the cell culture media (Figure 45F). Immunofluorescence analysis of neurons 4 hours after protein addition confirmed that less protein was applied (Figure 47). While large foci of MeCP2 protein that stained negative for NF were still apparent, they were fewer and there were a number of MeCP2 foci that co-stained for NF and appeared to be localised to intact neurons (arrows, Figure 47). This data suggests that a cleaner preparation of MeCP2 protein, applied in smaller amounts is able to transduce LUHMES-derived neurons while limiting cell death.

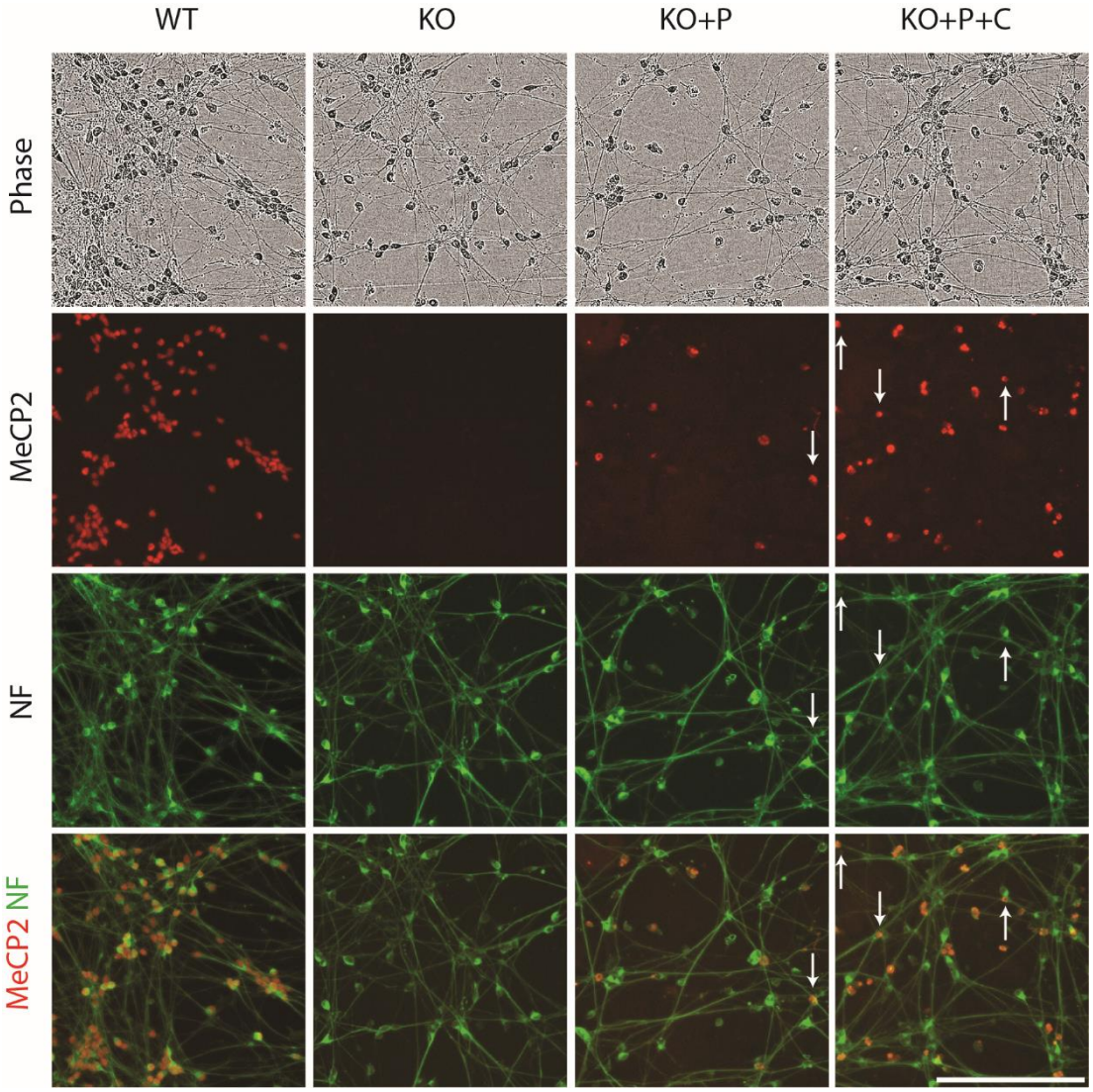


Figure 47. MeCP2 protein transduction in LUHMES-derived neurons using a new batch of protein.

Incucyte images of LUHMES-derived neurons taken on day 9, after 4 hours of exposure to either MeCP2 protein (P), chloroquine (C) or both (P+C). MeCP2 antibody is D4F3 (Cell Signalling). Scale bar is 100 µm.

Further analysis of neurons transduced with lower amounts of the new MeCP2 protein preparation will be necessary. In particular, high magnification images of neurons that are co-stained with DAPI will help to assess the fraction of neurons that have condensed or fragmented DAPI foci. Furthermore, Western blot analysis of neurons 24 hours after protein addition (Figure 45F) suggests that, as with the initial protein preparation, apoptotic pathways are being initiated. Thus immunofluorescence comparisons between 4 hours and 24 hours will be necessary to assess this. Overall, these experiments have demonstrated that MeCP2 protein, on its own or in combination with chloroquine, can transduce LUHMES-derived neurons to a certain degree.

7.3 Discussion

The ability to deliver proteins and other macromolecules directly to cells without the need for transfection techniques is challenging but extremely desired for therapeutic purposes. There has been a lot of work surrounding the use of so-called cell penetrating peptides (CPPs), as the ability to transduce cells simply by adding a small tag onto your protein of interest would make such an approach applicable to a wide variety of cells, tissues and protein molecules.

The discovery that Cre protein on its own is able to transduce eukaryotic cells, and that this activity is enhanced with a 6xhistidine-tag suggests that basic proteins might have an intrinsic ability to transduce cells without the need for CPPs (Lin *et al.*, 2004; Will *et al.*, 2002). Indeed, basic, polycationic peptides have been reported to be effective at crossing the plasma membrane, with polyarginine runs being the most efficient (Futaki *et al.*, 2001; Mi *et al.*, 2000; Mitchell *et al.*, 2000; Wender *et al.*, 2000). MeCP2 is another basic protein (Lewis *et al.*, 1992), which suggested that it may be able to transduce cells without the need for a CPP.

The experiments here have demonstrated that MeCP2 protein on its own can enter human, LUHMES-derived neurons and is highly suggestive that this protein subsequently enters the nuclei of these neurons. MeCP2 has an endogenous NLS which combined with its MBD aid the translocation of MeCP2 into nuclei (Lyst *et al.*, 2016; Nan *et al.*, 1996). Thus any MeCP2 protein that successfully crosses the plasma membrane would be expected to enter nuclei and bind to methylated DNA. While MeCP2 protein does enter LUHMES-derived neurons, the efficiency appears to be quite low, and the process itself appears to induce apoptosis. Performing immunofluorescence analysis and co-staining for apoptotic markers would help to confirm this, for example using the Annexin V-FITC/PI Apoptosis Detection Kit (BD Pharmingen) or TUNEL staining (Orii *et al.*, 2006).

The previous observation that chloroquine was necessary for MeCP2 transduction in MEFs was not true for LUHMES-derived neurons. Immunofluorescence analysis of MEFs (transduced with the new protein preparation at 1.25pg/cell) suggests that actually chloroquine has little effect on the ability of MeCP2 protein to enter MEFs (Supplementary Figure 14). Whether chloroquine was included or not, MeCP2 protein was observed attached to the plasma membrane, in the cytoplasm, and, on occasion, in the nuclei of MEFs (white arrows). The apparent immunoreactivity of the anti-neurofilament antibody in MEFs, while

useful for observing the cytoplasm, was a surprise. The positive staining could reflect non-specific antibody binding.

In order to understand the dynamics of protein translocation and subsequent neuronal death, live cell imaging studies would be extremely useful. A preparation of fluorescently tagged MeCP2 could be applied to neurons imaged in real-time to understand if dying neurons are transduced by the protein, or if neurons that take up the protein subsequently die. There is the possibility that tagging of MeCP2 with a large fluorescent molecule will affect its ability to transduce cells, particularly if the overall positive charge of the molecule is lost, so this would need to be assessed.

While the experiments here suggest that decreasing the amount of protein applied to cells results in decreased cell death, Western blots after 24 hours were suggestive of apoptotic pathways. It would be interesting to decrease the amount of MeCP2 protein further in an attempt to avoid cell death pathways. Live cell imaging may even help to observe two neurons side by side that take up different amounts of protein, and thus perhaps survive for different lengths of time. Such studies could be combined with tools that fluoresce upon apoptotic triggers, such as acridine orange (Galluzzi *et al.*, 2009). In addition to this, small amounts of MeCP2 could be added by multiple applications over a prolonged period of time in order to avoid cell death. Such an experiment would be reminiscent of the procedure used for the MeCP2 reversal experiments in mice where instant reactivation of the inactive *MeCP2* allele often resulted in toxicity (Guy *et al.*, 2007).

Immunofluorescence analysis indicated that MeCP2 loading is uneven among all neurons in the population. Some neurons take up extremely high quantities of protein while others remain MeCP2 null. Perhaps mixing MeCP2 with a CPP will encourage a more even loading that would ultimately reduce the population cell death levels. Mixing a sgRNA molecule with a polyarginine CPP successfully transduced the negatively charged sgRNA into cells, and while the purpose of the CPP was to alter the electrostatic properties of the sgRNA, perhaps interactions between MeCP2 and the CPP will aid protein delivery (Ramakrishna *et al.*, 2014). Alternatively MeCP2 could be tagged with a CPP, or the iTOP method which uses NaCl and a propanebetaine to encourage pinocytic uptake of protein (D'Astolfo *et al.*, 2015) could be used to encourage more even loading and reduced levels of cell death.

While these experiments begin exploration of protein transduction therapy for Rett syndrome, adapting the process for *in vivo* work will not be trivial. Direct injection into sites in the brain, as is used for AAV delivery in mouse models (Gadalla *et al.*, 2017) and in clinical trials of stem cell injections (Steinberg *et al.*, 2016), or intrathecal injections (Calias *et al.*, 2014) would give the best chance of efficient delivery. The isoelectric point (pI) of MeCP2 is very high (~10) (Lombardi *et al.*, 2015) suggesting that systemic injections of MeCP2 protein would likely result in clearance of the protein resulting in low therapeutic efficacies, as has been observed in the minipig with monoclonal antibodies (Zheng *et al.*, 2012). Recently, pre-assembled, CPP-tagged Cas9 RNPs have been stereotaxically injected into four brain regions in the mouse brain and *in vivo* CRISPR-induced DNA editing was observed (Staahl *et al.*, 2017). This suggests that protein delivery to the brain and subsequent cellular protein transduction is possible *in vivo*. The authors observed a larger spread of protein transduction throughout the injected brain region when a higher concentration of protein was injected, however the transduction was still localised around the injection site, thus the therapeutic potential of such a delivery method is still questionable (Staahl *et al.*, 2017).

More experiments are needed to determine if a lower dosage of protein will circumvent the cell death that has been observed in this study. Future assessment of other parameters, such as mixing with CPPs or iTOP, may also prove to be beneficial for elucidating an efficient and reproducible protein delivery method. LUHMES KO cells can be expanded to large quantities as precursors and differentiated in a high throughput manner in 24-well plates, thus simultaneous testing of multiple different parameters will be possible for future experiments.

Chapter 8

8. Discussion and future perspectives

8.1 Mutant LUHMES cell lines as model systems for studying MeCP2 and Rett syndrome

Overall, the data presented in this study suggests that the LUHMES cell line and its derived neurons will be a useful model system for future studies into MeCP2 and Rett syndrome. The cells are easy to handle, can be genetically manipulated and rapidly differentiate into a homogeneous population of mature neurons. In addition, experiments using MeCP2 mutant neurons reveal robust changes to the transcriptome and future studies using these cells will help to elucidate the molecular mechanisms behind these alterations. In particular, studies such as 4sU-sequencing will be easier and cheaper to perform in tissue culture model systems compared to rodent models due to the ability to produce large quantities of starting material allowing for numerous biological and technical replicates to generate reliable data.

One negative to LUHMES-derived neurons in terms of studying the function of MeCP2 is the low abundance of mCA that exists in the genome. While LUHMES-derived neurons have 14% mCH/mC, the amount *in vivo* increases during development reaching ~53% of all methylated cytosine residues in the adult human brain (Lister *et al.*, 2013). Recent analysis comparing neurons derived from young and old fibroblasts demonstrates an inherent difference in the age of the resulting neurons as judged by epigenetic signatures, miRNA profiles and telomere length, among other assays (Huh *et al.*, 2016). This suggests that global mCH levels could be another age-related signature that is inherent to the cell, and as LUHMES cells were derived from 8-week old fetal tissue their derived neurons may never be able to generate high levels of mCH on par with the adult brain. Artificially increasing mCH levels by overexpression of DNMT3A could be used to boost mCA amounts in order

to see if MeCP2 dynamics are altered and transcriptional outputs differ. Nevertheless, the MeCP2-regulated transcriptional changes observed in this study, where only mCG was studied, revealed similar mechanisms as mice studies that assess mCA and mCG (Kinde *et al.*, 2016; Lagger *et al.*, 2017). This suggests that boosting mCA levels in LUHMES cells will not be necessary for understanding the mechanism of MeCP2 function in human neurons.

A second disadvantage of LUHMES-derived neurons is the inability to differentiate the precursors into alternative neuronal lineages. LUHMES precursor cells appear to be committed to a dopaminergic lineage (this study and (Scholz *et al.*, 2011)). Removal of GDNF and cAMP from the differentiation media results in decreased levels of *TH, RET* and *DRD2* in LUHMES-derived neurons as well as reduced dopamine levels. But the unaltered expression level of six other dopaminergic markers and four other neurotransmitter markers suggests that a change in neurotransmitter type does not occur under these culture conditions (Scholz *et al.*, 2011). While MeCP2 likely has the same function in different neuronal subtypes (to bind to methylated DNA and repress transcription), the output of this mechanism of action is likely to be different in different neurons and thus the phenotype when MeCP2 function is perturbed will vary. Evidence for this is already apparent with chromatin compaction only being observed in hippocampal CA1 pyramidal neurons and hippocampal dentate granule cells but not in cerebellar granule cells upon MeCP2 deletion (Linhoff *et al.*, 2015). Likewise the effect of MeCP2 overexpression on neurite complexity is the opposite in hippocampal cultures to that in cortical cultures (Jugloff *et al.*, 2005; Zhou *et al.*, 2006). Thus, for detailed phenotypic analysis of specific neuronal populations, hiPSC-derived neurons and *in vivo* studies in rodents would be a more appropriate model than LUHMES-derived neurons.

The transcriptome map produced in this study of WT LUHMES-derived neurons could be complemented with similar analysis of LUHMES neuronal precursors and the two datasets then compared to published datasets of human brain development using the CoNTeXT framework. This would be useful for assessing the human developmental time point that LUHMES-derived neurons correspond to, and to observe which neuronal gene expression networks are preserved in the tissue-culture differentiation system (Stein *et al.*, 2014). This framework was built by comparing the transcriptome of multiple tissue-culture differentiation systems, including hiPSCs and primary neural progenitors (phNPCs), to *in vivo* neuronal development from 4 weeks post-conception to 40 years of age. The analysis

found that hiPSCs produce neurons that have a more immature transcriptome compared to phNPCs and that neither phNPCs nor hiPSCs could be encouraged to differentiate past fetal-stage neurons even if left in culture for months (Stein *et al.*, 2014). In addition to CoNTeXT, the LUHMES transcriptome could be compared to that from human dopaminergic iPSC-derived neurons and laser-captured human dopaminergic neurons, which have already been compared transcriptionally to one another (Sandor *et al.*, 2017). Thus, one could use such analyses to more accurately determine the stage of development ("maturity") that LUHMES-derived neurons correspond to and compare their transcriptional profile to *in vivo* dopaminergic neurons from an adult human brain. Furthermore, comparison of the active gene networks in LUHMES-derived neurons to *in vivo* brain development may help to determine the expected phenotypic outputs in tissue culture.

Future avenues of research using LUHMES-derived neurons could also explore co-culture with astrocytes or the development of organoids (Ardhanareeswaran *et al.*, 2017). Such systems are known to produce neurons that are more mature, and perhaps more responsive to external cues thus analysis of MeCP2-induced phenotypes in such a culture system might better represent the *in vivo* situation in RTT patients (Johnson *et al.*, 2007; Kuijlaars *et al.*, 2016; Odawara *et al.*, 2014; Tang *et al.*, 2013).

In summary, I believe that LUHMES-derived neurons will be a valuable tool for assessing the molecular defects that occur upon MeCP2 mutation. In addition, the demonstration in this study that neurons can be differentiated in a high-throughput manner in 96-well plates highlights the potential utility of these neurons for drug screening purposes. The ease of differentiation, efficient genetic manipulation and appearance of strong phenotypes suggests that LUHMES-derived neurons are an attractive model system to complement *in vivo* mouse studies.

8.2 Transcriptional profiling in Rett syndrome

One particular area of research that the LUHMES cell line will likely contribute to is transcriptome analysis in MeCP2 mutant neurons. There have now been numerous studies assessing the transcriptome of MeCP2 mutant neuronal samples, both in tissue culture (Li *et al.*, 2013c), in mouse brain regions (Ben-Shachar *et al.*, 2009; Chahrour *et al.*, 2008; Chen *et al.*, 2015; Gabel *et al.*, 2015; Jordan *et al.*, 2007; Kinde *et al.*, 2016; Rube *et al.*, 2016) and in patient cortex (Colantuoni *et al.*, 2001). While early microarray studies hypothesised a role

for MeCP2 in transcriptional activation (Ben-Shachar *et al.*, 2009; Chahrour *et al.*, 2008; Li *et al.*, 2013c), recent RNA-sequencing analysis combined with high coverage ChIP-sequencing and methylation mapping have confirmed the association of MeCP2 with gene repression that was discovered in early repressor assays (Gabel *et al.*, 2015; Lagger *et al.*, 2017; Nan *et al.*, 1997). The RNA-sequencing analysis in LUHMES-derived neurons from this study further confirms the role of MeCP2 in gene repression and additionally suggests that MeCP2 functions via gene-body methylation. While gene-body methylation has been suggested as being important for MeCP2-mediated gene repression in a prior mouse study (Gabel *et al.*, 2015), large domains of methylation have also been implicated as the mode of MeCP2 repression (Lagger *et al.*, 2017). In the LUHMES dataset, gene body methylation levels and surrounding methylation levels are very highly correlated, thus distinguishing the contributions of each to gene repression will be difficult (J Cholewa + RR Shah *et al*, unpublished results). By analysis of short genes within large methylation domains, Kinde and colleagues determined an individual genes methylation as being more critical for repression than being situated in a methylated domain (Kinde *et al.*, 2016). Similar bioinformatics approaches should be applied to the LUHMES datasets in order to assess the importance of gene body and flanking region methylation on MeCP2-mediated gene regulation.

While steady-state RNA-sequencing datasets have been vital in determining the role of MeCP2 in gene expression, future studies should look at dynamically changing and nascent gene transcription. In its native environment a neuron is rarely in a steady-state, but is in constant communication with surrounding neurons as part of an active network. Assessment of the transcriptional profile of WT and mutant neurons before and after KCl-induced depolarisation might reveal stark defects in the ability of MeCP2 mutant neurons to respond to depolarisation signals. In particular qPCR experiments assessing the activation of early response genes such as *JUN* and *FOS* immediately after depolarisation and the late response genes such as *BDNF* and *GPR22* at later time points might be insightful. Indeed, excessive induction of *Junb,* and *Arc* (but not *Fos*) has been observed in *Mecp2*^{T308} striatum after cocaine administration (Su *et al.*, 2012), and activity-dependent phosphorylation of MeCP2 has been reported to allow for increased *BDNF* transcription (Chen *et al.*, 2003; Zhou *et al.*, 2006) and increased *Npas4* transcription (Ebert *et al.*, 2013). These studies suggest that MeCP2 functions to repress activity-regulated genes, and that loss of MeCP2 binding to genes (by genetic mutation or by phosphorylation) results in increased expression of these genes. In contrast to these studies, experiments in *MECP2*-null human ESC-derived neurons

found reduced expression of immediate-early response genes *ARC*, *FOS*, *NPAS4* and *BDNF* in a basal state, and a lack of induction of these genes after KCl-induced depolarisation (Li *et al.*, 2013c). Similar RNA-sequencing and qPCR experiments in the mutant LUHMES-derived neurons generated in this study may reveal differences in the expression of immediate-early genes, as well as global transcriptional profiles after KCl-induced depolarisation.

In order to fully understand the effect MeCP2 binding has on gene expression, nascent RNA-sequencing technologies should be employed such as 4sU-sequencing (Cleary *et al.*, 2005). In particular treatment of neurons with DRB to first inhibit all transcription, combined with DRB wash-out and 4sU treatment will provide accurate datasets of nascent transcription and elongation rates in WT and MeCP2 mutant samples (Fuchs *et al.*, 2016). Integration of elongation rates and MeCP2 ChIP-seq data will help to determine the effect of MeCP2 binding (perhaps throughout gene bodies) on the Pol II elongation rate in genes. Combined with genome-wide mapping of RNAP II initiation, elongation and termination (as judged by its various phosphorylation states) these datasets will help towards an understanding of the mechanism of MeCP2-mediated gene repression.

While RNA is an essential intermediate in gene expression, the proximal cause of phenotypic variation is likely to be at the level of protein. Therefore, in addition to more detailed assessments of the transcriptome, future studies could focus on the proteome in mutant neurons. The finding in this study that total RNA is reduced in both MeCP2 KO and R306C neurons suggests that total protein levels will be decreased as well. This will be especially pertinent if, as hypothesised, rRNAs are reduced in mutant LUHMES-derived neurons. Among the RNA species that Li and colleagues found reduced in *MECP2*-null hESC-derived neurons, were a number of mRNAs coding for ribosomal proteins as well as rRNAs (Li *et al.*, 2013c). This suggested that mutant neurons might have an inability to efficiently translate proteins and in line with this, incorporation of radiolabelled cysteine and methionine amino acids was found to be reduced by 25% (Li *et al.*, 2013c). Apart from assessing the importance of the proteome for RTT pathologies and for identifying potential therapeutic avenues, proteome analysis may reveal biomarkers that can be used when testing compounds in drug screens.

8.3 Concluding remarks

This project set out to assess the utility of the LUHMES cell line as a new model system for studying MeCP2 and Rett syndrome. The project begun with some basic characterisations of the cell line and its derived neurons and found that LUHMES cells differentiate quickly and efficiently into a homogeneous population of neurons. As to the maturity of these neurons, markers such as NeuN, MAP2, Neurofilament, TH, DAT and PSD-95 were all found to be expressed suggesting that mature neurons were produced within 9 days. As to how this level of maturity compares to neurons in an *in vivo* context is unknown and direct transcriptome comparisons between LUHMES-derived neurons and brain-derived neurons during development should be performed to assess this further.

The demonstration that efficient and reproducible transfection, single cell cloning and genetic engineering techniques can be performed in LUHMES cells demonstrates the ease of use of this cell line. The ability to manipulate LUHMES cells will be useful for modelling disease-causing Rett syndrome mutations, for cellular perturbation to probe MeCP2-related molecular pathways, and will be applicable for uses outside the MeCP2/Rett syndrome field.

Assessment of MeCP2-null LUHMES-derived neurons found no decreases in neurite length or nuclear volume. This is in contrast to published reports of similar studies using alternative model systems and is perhaps indicative of a limitation of LUHMES cells with respect to mirroring disease pathology. Being a population of neurons alone, as opposed to the mixed cellular environment in brain slices and iPSC systems, perhaps limits the ability of LUHMES-derived neurons to phenocopy neurons in an *in vivo* context. This would suggest that the environment a neuron is situated in can affect the phenotypic outcome. As such, LUHMES-derived neurons are useful for studying the inherent phenotypes within dopaminergic neurons, but will not be useful for recapitulating the endogenous disease state.

Molecular analyses, on the other hand, can readily be probed in LUHMES-derived neurons. The ability to generate large quantities of neurons in a short space of time using a simple differentiation procedure means that LUHMES cells are a great candidate for a plethora of molecular and biochemical studies. This is illustrated by the number of biochemical perturbations that were found in mutant LUHMES-derived neurons in this study. The approximate 10% decrease in total RNA, and the gene-body methylation dependent gene expression changes were reproducible between cell lines and among numerous MeCP2-

mutant samples. The strong relationship between transcription and gene-body methylation, which shows a reciprocal effect between MeCP2 null and MeCP2 overexpression cell lines, suggests that the primary role of MeCP2 in human neurons is to repress transcription, and that MeCP2 achieves this goal by binding to sites of methylation within the transcriptional unit of a gene.

While this is a significant step forward, a full understanding of how MeCP2 achieves this gene repression is still unknown. Is the recruitment of the entire NCoR/SMRT complex critical for this constraint on gene expression or is TBL1/TBLR1 recruitment alone sufficient? Does MeCP2 hinder RNA polymerase II elongation through the gene body? Is the chromatin environment altered in MeCP2 mutant samples, and if so, is this a primary or secondary or even compensatory effect? Is the total RNA decrease a primary effect from MeCP2 mutation or secondary, due to "sub-optimal" neurons? Future directions to assess the mechanism of MeCP2-mediated gene repression will require genome-wide analysis of nascent transcription, for example by 4sU-sequencing. Complementary to this will be ChIP-sequencing of total and phosphorylated RNA polymerase II to assess the ability of RNA polymerase II to initiate and elongate in mutant samples. The decrease in total RNA observed in MeCP2-mutant samples is perplexing for a protein that is proposed to repress transcription, not activate it. Further studies to evaluate the importance of this finding will involve analysing mice before the onset of symptoms. If young mice have decreased total RNA before symptoms have emerged, this might suggest a causal role of reduced RNA levels in Rett-like symptoms. Other experiments will involve assessing rRNA transcription by Northen blot analysis, measurement of nucleolar volumes (the sites of rRNA transcription, processing and ribosome assembly), and measurements of the number and volume of UBF compartments (the sites of rRNA transcription).

Since the discovery of MeCP2 25 years ago and its link to Rett syndrome 18 years ago huge progress has been made in multiple areas; investigating the function of MeCP2, identifying that Rett syndrome is a reversible disorder, and working towards trials of therapeutic strategies. There is still a long way to go, however, for a full molecular understanding of how MeCP2 functions and precisely how this is perturbed in Rett syndrome. LUHMES-derived neurons will be extremely useful for probing these molecular questions and may even prove to be valuable in the search for therapeutics for Rett syndrome. As the field moves forward the use of multiple model systems, and in particular studies in human neurons, will help to uncover a full molecular aetiology of and, ultimately, a cure for Rett syndrome.

Chapter 9

9. Supplementary Information

9.1 Tissue culture RTT model systems

System	Gender	Mutation	Reference
Mouse ESCs	Male	Exon 3 + 4 deletion	(Yazdani *et al.*, 2012)
Human ECSs	Male + Female	Exon 3 deletion	(Li *et al.*, 2013c)
RTT iPSCs	Female	T158M R306C Q244X 1155del32	(Marchetto *et al.*, 2010)
RTT iPSCs	Female	T158M V247X R306C	(Ananiev *et al.*, 2011)
RTT iPSCs	Female	Exons 3 + 4 deletion	(Cheung *et al.*, 2011)
RTT iPSCs	Female	T158M Q244X E235fs R306C X487W	(Kim *et al.*, 2011)
Mouse iPSCs	Female	T308X	(Farra *et al.*, 2012)
MeCP2dup iPSCs	Male	300 kb 500 kb 15 Mb	(Nageshappa *et al.*, 2015)
RTT iPSCs	Male Female	Q83X N126I	(Tang *et al.*, 2016)
RTT iPSCs	Male Female	Q83X N126I	(Zhang *et al.*, 2016)
RTT iPSCs	Female	1155del32	(Hunihan *et al.*, 2017)
RTT/mouse iPSCs	Female	R306C/T308X	(Hotta *et al.*, 2009)
RTT iPSCs	Female	1155del32 T158M	(Pomp *et al.*, 2011)
Human ESCs RTT iPSCs	Female Female	T158M KO 517bp deletion V247fs	(Bu *et al.*, 2017)

Supplementary Table 1. List of tissue-culture model systems for studying MeCP2 and RTT.

9.2 Neuronal differentiation defects in RTT model systems

System	Difference	Technique	Reference
T158M-iPSC derived neurons	50% decrease in number of cells that are TuJ positive*	Immunofluorescence of TuJ in neurons.	(Kim *et al.*, 2011)
	Decrease in *TuJ**, *PAX6**, *SCN1A*, *SCN1B*	qRT-PCR analysis	
MeCP2$^{lox/y}$/CreER injected with tamoxifen at 15 or 34 weeks to ablate MeCP2, age of mice at experiment is "within days" of tamoxifen treatment.	Syn1 to 25%. VGLUT1 to 40%. NMDAR2A to 40%. SHANK1 + SHANK2 to 50%. No change in SHANK3, CAMKIIa or b, GABABR2, GluR2/3, PSD93, SYT1.	Western blot. 3 mice per group. No replicates of Western blots? B-actin used as loading control. Odyssey scanner used to scan membranes and measure band intensities.	(Du *et al.*, 2016)
Brain slices of 11 patients with classical RTT + 12 age-matched controls	Tendency for MAP2 staining to be decreased* but only significantly different in 1/3 of brain regions. Tendency for NF staining to be increased and decreased, but only significantly different in 1/3 of brain regions.	Brain slices, fixed and stained for MAP2 and NF. Layers II-III and V-VI in somatosensory / motor cortex assessed. Both somatic and dendritic staining assessed.	(Kaufmann *et al.*, 2000)
Nasal biopsies from 29 RTT patients (8 of which had no MeCP2 mutation)	Increased staining of NST (immature olfactory neuron marker and increased staining of OMP (mature olfactory neuron marker)	Nasal biopsies, fixed and stained for NST and OMP. Control was 31 normal male and females.	(Gabriele V. Ronnett *et al.*, 2003)

Supplementary Table 2. Neuronal differentiation defects in RTT model systems.

RTT model systems that have studied the maturation of neurons containing MeCP2 mutations and found defects in neuronal differentiation as judged by the expression of various marker proteins.

* No change in TuJ or PAX6 as determined by immunofluorescence in R294X-iPSC derived neurons (Ananiev *et al.*, 2011).

* No change in MAP2 in male mice that express MeCP2 that is truncated at residue 308 (Shahbazian *et al.*, 2002b).

9.3 Nuclear size defects in RTT model systems

System	Difference	Technique	Reference
R294X-iPSC derived neurons	9% and 17% 2493 ± 24 AU → 2061 ± 18 or 2273 ±23 or 2267 ±50 AU.	DAPI staining circled in Photoshop and measured in Photoshop. 1 WT cell line, 3 iPSC clones from the same patient. N=378, 490, 279, 90	Ananiev *et al*. 2011
Mecp2$^{-/y}$ mESC derived neurons	25-30% 27 µm^2 → 20 µm^2 (DIV8) 32 µm^2 → 24 µm^2 (DIV26) No difference at 3 DIV	Hoescht staining of nuclei, ImageJ calculates size. N=131+184 for DIV8. N=49+61 for DIV26.	Yazdani *et al*. 2012
Mecp2$^{-/y}$ mouse brain slices, 1-month old CA3 region of the hippocampus	138.2 ± 2.1 µm^2 → 98.7 ± 1.6 µm^2 No change in size of astrocytes	Hoescht staining of nuclei, ImageJ calculates size. N = 56 WT and 92 KO.	
KO hESC-derived neurons (deletion of exon 3)	25% 100 µm^2 → 75 µm^2	DAPI staining of nuclei. No methods given. No details about replicates or numbers.	Li *et al*., 2013
MeCP2$^{-/+}$ mouse brain slices, 5 month old, hippocampal CA1 neurons	5% 155 ± 19 µm^3 147 ± 18 µm^3	ChromATin. N=51 for WT and 55 for null. From 3 mice.	Linhoff *et al*., 2015
MeCP2$^{-/y}$ mouse brain slices	33%	DAPI staining. No cut-off for fragmented DAPI or doublet/triple nuclei.	Greenberg, word of mouth
MeCP2$^{lox/y}$*/CreER* injected with tamoxifen at 15 weeks to ablate MeCP2, age of mice at experiment probably 17 weeks.	15% 175 µm^2 to 150 µm^2	Hippocampus CA1, DAPI colocalised with NeuN, ImageJ calculates. Three mice per group, 20 neurons per animal.	Du *et al*., 2016

Supplementary Table 3. Nuclear size defects in RTT model systems.

RTT model systems that have assessed the nuclear size or volume of MeCP2-mutant samples and found a decrease compared to WT controls.

* No change in nuclear size in olfactory neuroepithelium neurons from *MeCP2*$^{-/y}$ mouse (Rube *et al*., 2016)

9.4 Neurite complexity defects in RTT model systems

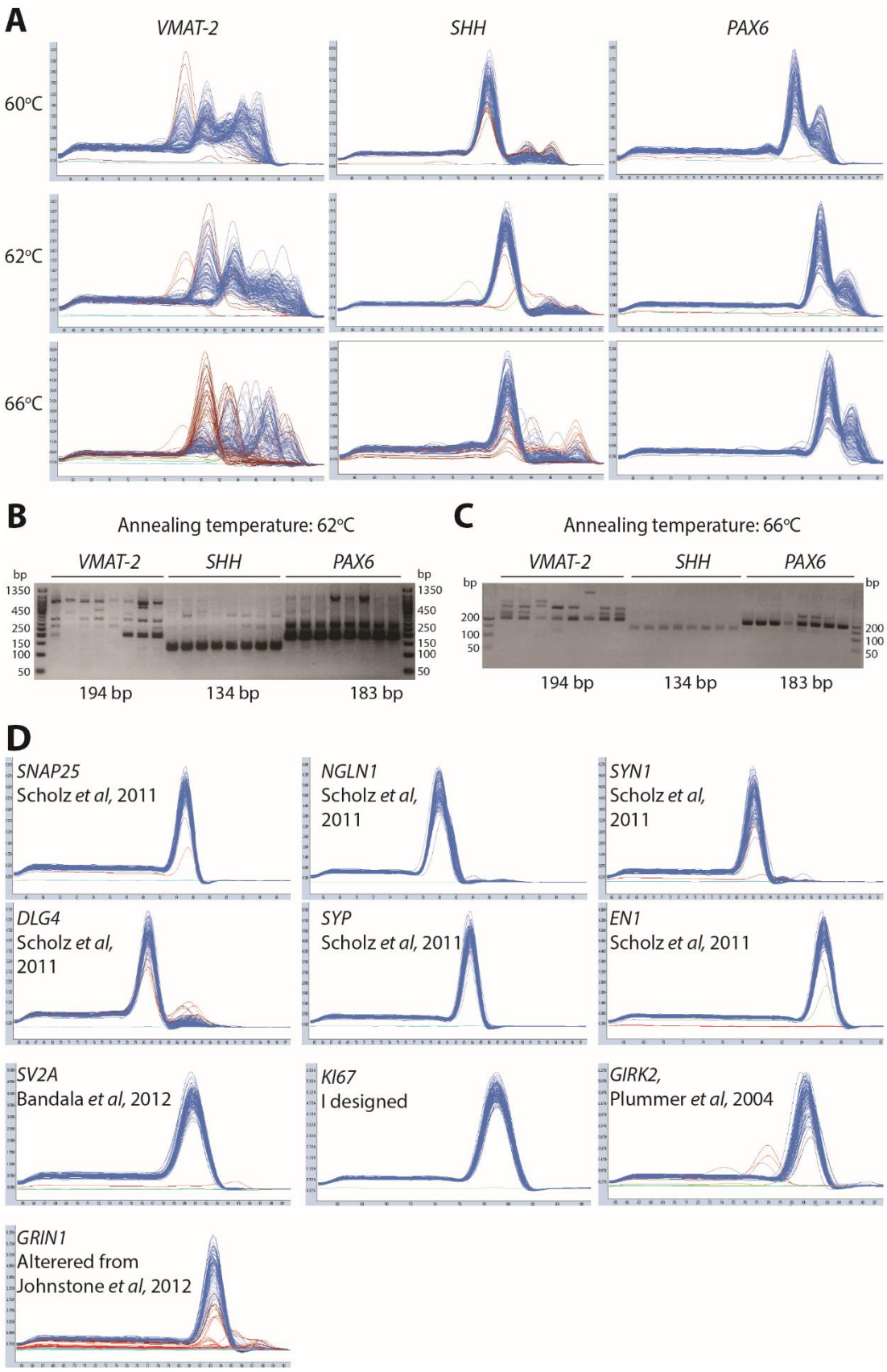
System	Difference	Technique	Reference
KO hESCs- derived neurons (de letion of exon 3)	Number of intersections decreased by about 1/3 at all distances along the neurite.	Scholl analysis using synapsin-GFP. Measured using ImageJ. No details about replicates or numbers.	(Li *et al.*, 2013c)
MeCP$^{-/y}$ mice: Bird and Jaenisch alleles	Dendrites: Decreased spine density, decreased spine area, increased number of 2.5 µm that have no spines, increased frequency spine swelling, trend for decreased spine widths, trend for increased spine neck length Axons: Increased density of axons, decreased alignment of axons	Brain slices. Injected with Lucifer yellow to observe dendrites or with carbocyanine to image axons. Motor cortex, CA1 region of hippocampus and fascia dentate were examined. For all measurements, n = 3 mice.	(Belichenko *et al.*, 2009)
MeCP2$^{lox/y}$*/CreER* injected with tamoxifen at 15 or 20 weeks to ablate MeCP2, age of mice at experiment "within days".	Dendritic length: 350→150 µm (50%). Number of nodes: 12→5. Spine density: 110→60 spines/100 µm. Number of crossings (scholl) decreased by ½ at peak and reduced at all distances.	Golgi impregnation, 3D neuronal reconstruction using camera lucida device, analysed by neurolucida explorer software for spine density, branches and Sholl analysis. CA1 pyramidal neurons.	(Du *et al.*, 2016)
MeCP2$^{lox/y}$*/CreER* injected with tamoxifen at 10 weeks to ablate MeCP2, age of mice at experiment 14 weeks.	Dendritic length: 25% reduction. No change in spine density (small change at 18 weeks).		
Brain sections of RTT patients, ages 2.9 – 35 years old.	Decreased branches in some, but not all brain regions tested.	Golgi impregnation, camera lucida drawings of pyramidal neurons in the cortex, subiculum and CA1 of the hippocampus. Sholl analysis. N = 10 neurons per sample, 16 patients.	(Dawna Armstrong *et al.*, 1995)

Supplementary Table 4. Neurite complexity defects in RTT model systems.

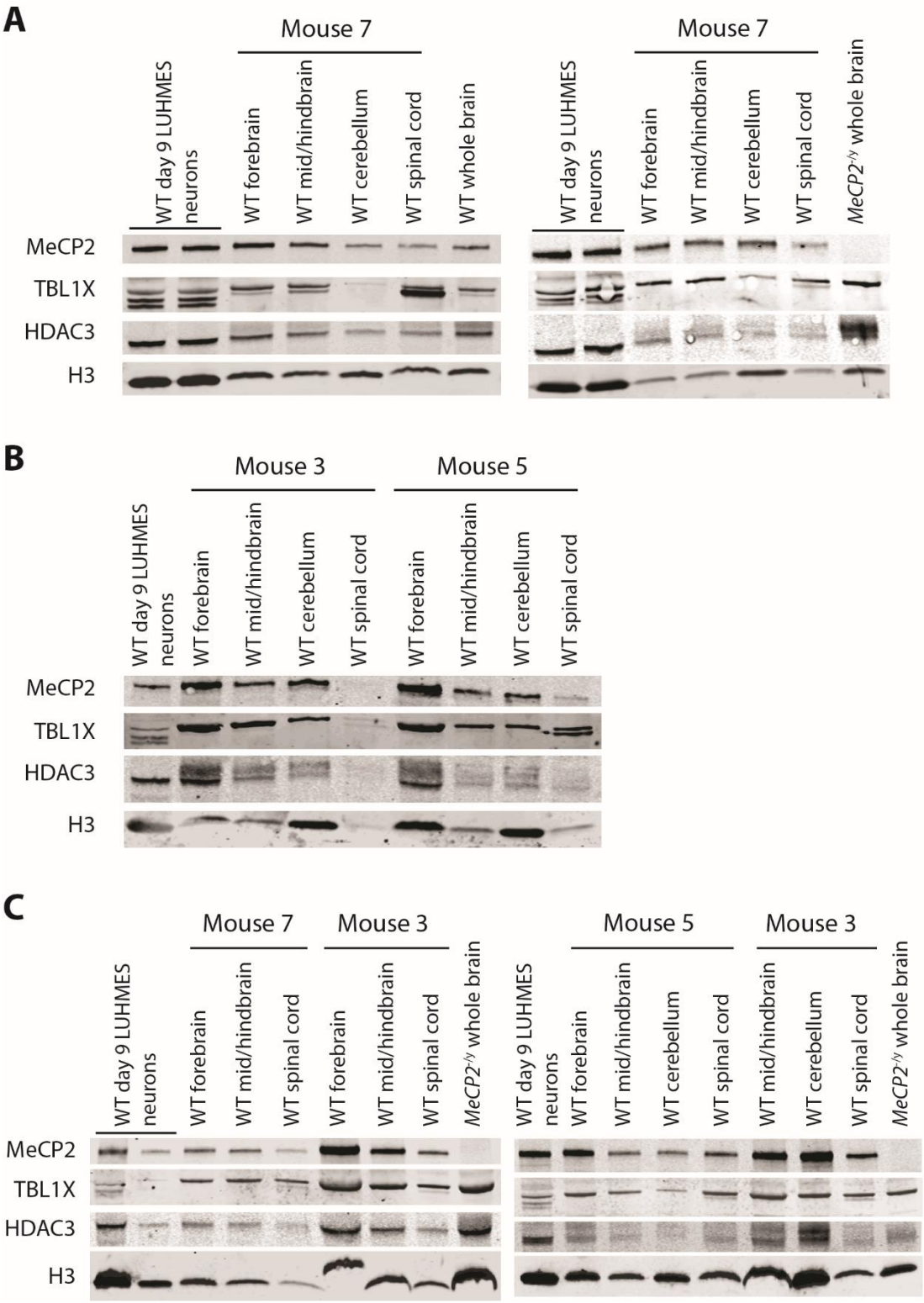
Five model systems that have been assessed for neurite morphology and found defects in the MeCP2-mutant neurons compared to WT controls.

* No defect in neuronal morphology was observed in the cortex of male mice expressing MeCP2 that is truncated at residue 308 as judged by Golgi staining, camera lucida drawings and Scholl analysis. They did, however, by transmission electron microscopy of the CA1 region observe a reduction in the number of larger post-synaptic density neurons but no alterations in the total number of synapses was observed (Moretti *et al.*, 2006).

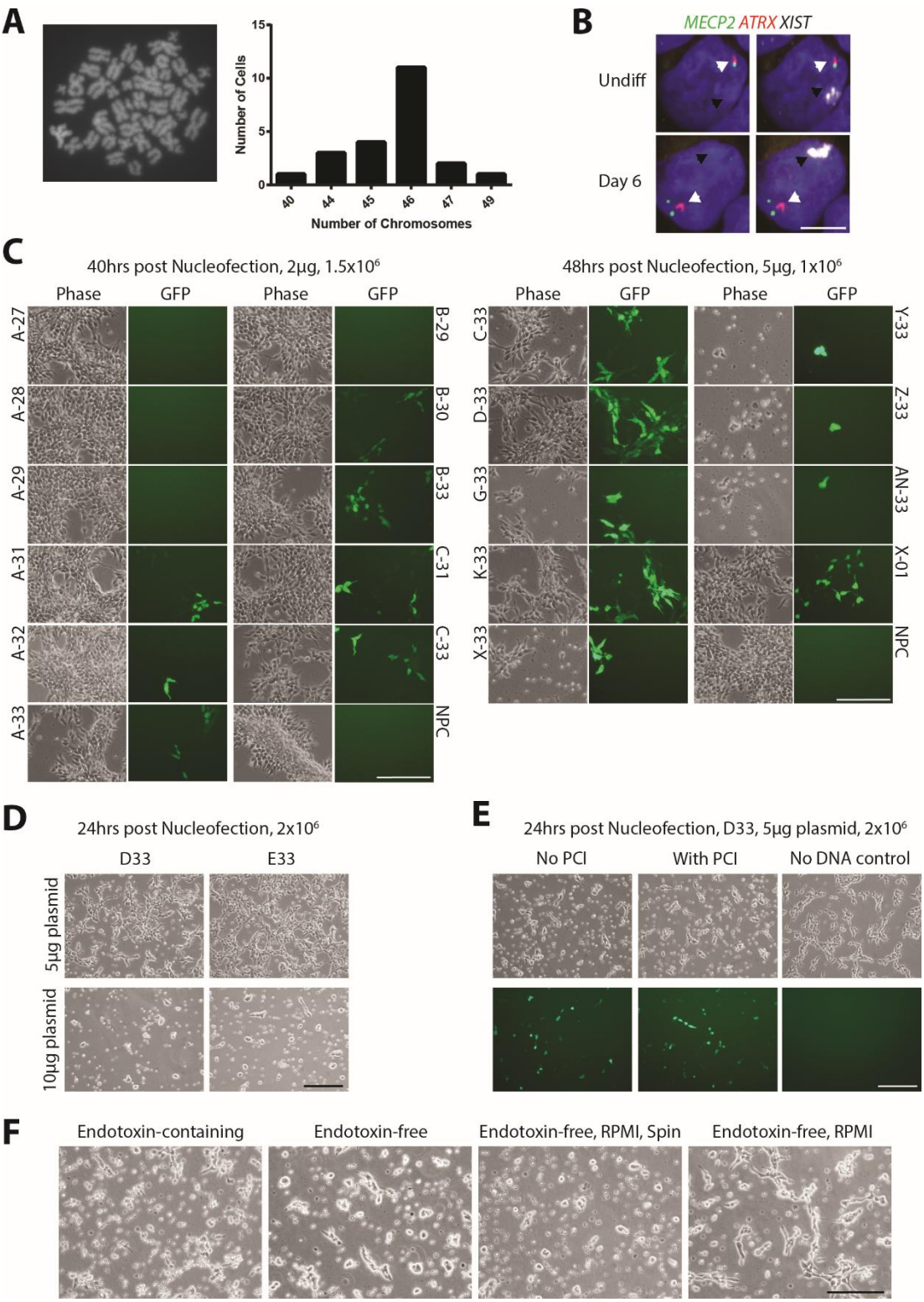
9.5 Supplementary Figures



Supplementary Figure 1. Analysis of new primer pairs for qPCR analysis.



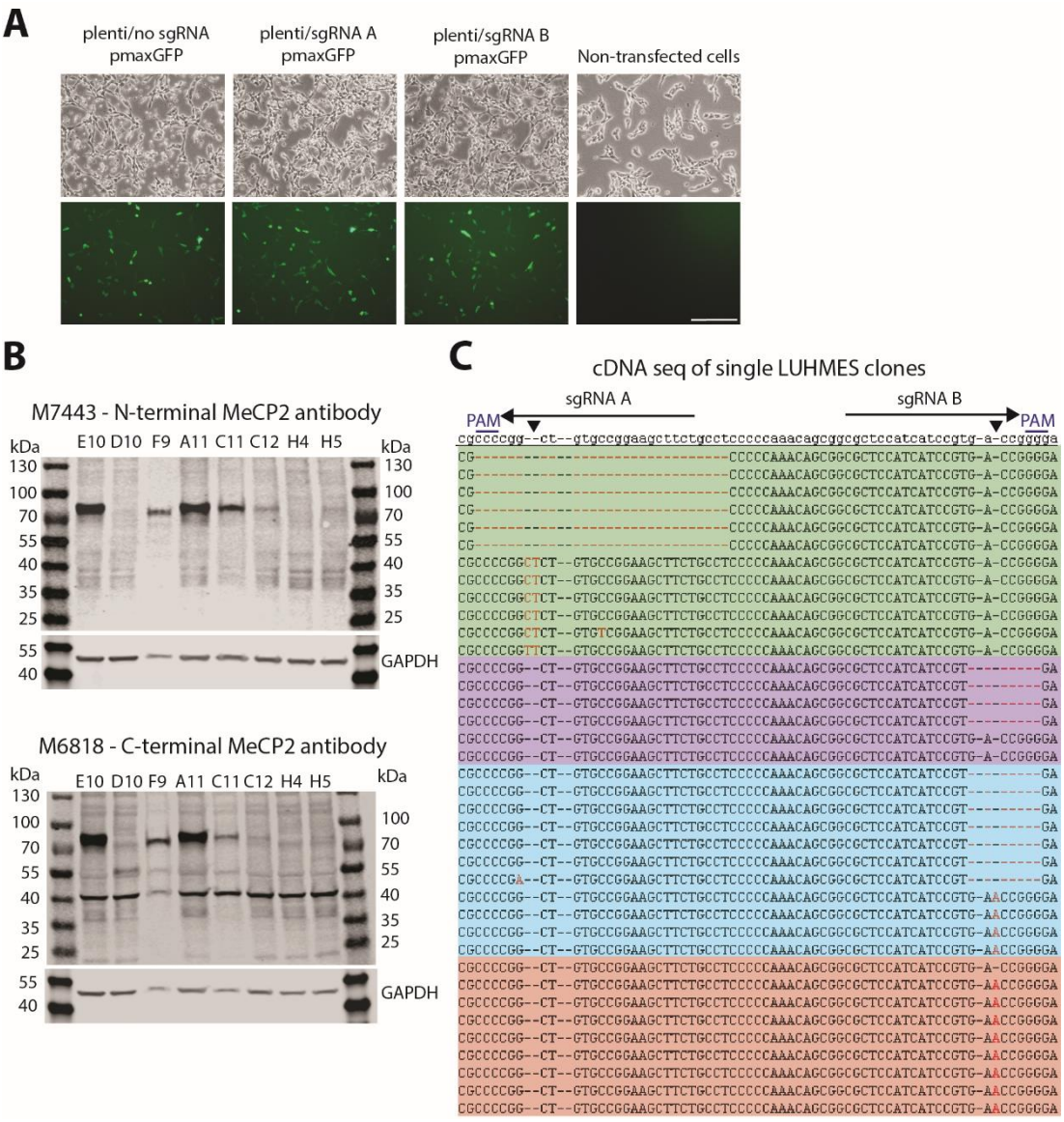
Supplementary Figure 2. Replicate Western blots of MeCP2, HDAC3 and TBL1X in LUHMES-derived neurons and mouse brain regions.



Supplementary Figure 3. Optimisation of Nucleofection in LUHMES cells.

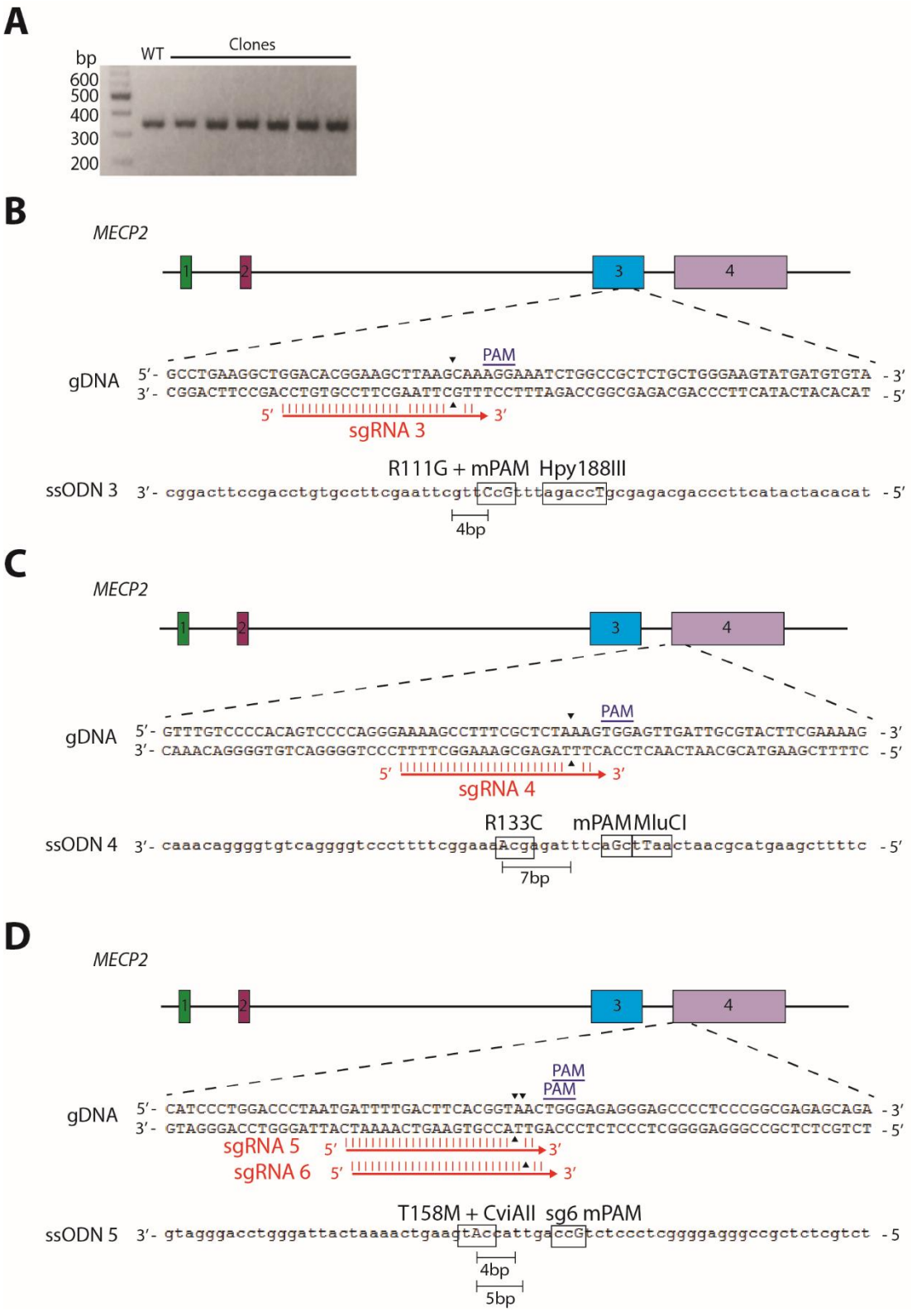
(A) Karyotyping of LUHMES pre-neuronal cells confirms a normal ploidy. **(B)** RNA FISH performed by Dr Ronan Chaligne of these female cells demonstrates that X inactivation is already established in the pre-neuronal cells. Scale bar is 5 µm. **(C)** Trials of various electroporation programs using the Basic Nucleofector kit for primary neurons. **(D)** Comparison of different plasmid amounts on cell viability. **(E)** Comparison of phenol:chloroform:isoamylalcohol plasmid purification on cell viability.

(F) Comparison of cell viability when endotoxin-free plasmids are used, when the cells are incubated at 37°C in RPMI for 5 minutes (RPMI) and when they are centrifuged before plating (spin). All scale bars are 200 µm.



Supplementary Figure 4. Analysis of MeCP2 KO cell lines.

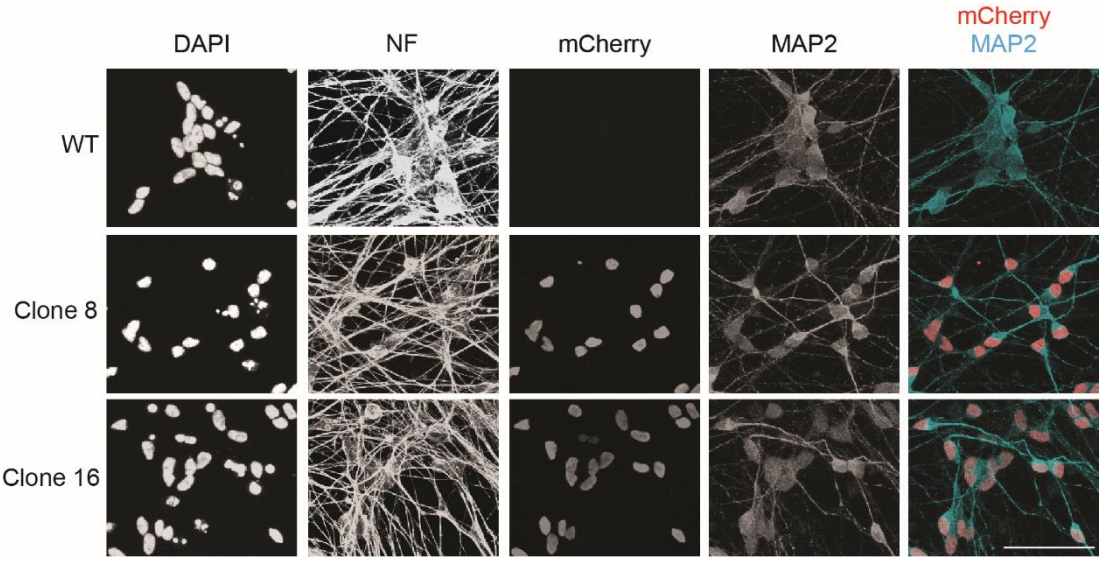
(A) Representative phase contrast and GFP fluorescence images of LUHMES cells 25 hours after Nucleofection. Scale bar is 200 µm. **(B)** Immunoblot detection of MeCP2 protein in single cell clones derived from the plentiCRISPR targeting experiment using an N-terminal and a C-terminal antibody. GAPDH is probed for a loading control. Note: E10 = WTC, D10 = KO1 and H4 = KO2 in Fig 2. **(C)** Sequencing of cDNA from single LUHMES clones identifies mixed cell populations by the presence of multiple cDNA alleles per clone. Each colour block denotes an individual clone.



Supplementary Figure 5. Optimisation of CRISPR-mediated point mutation knock-in of Rett syndrome causing missense mutations.

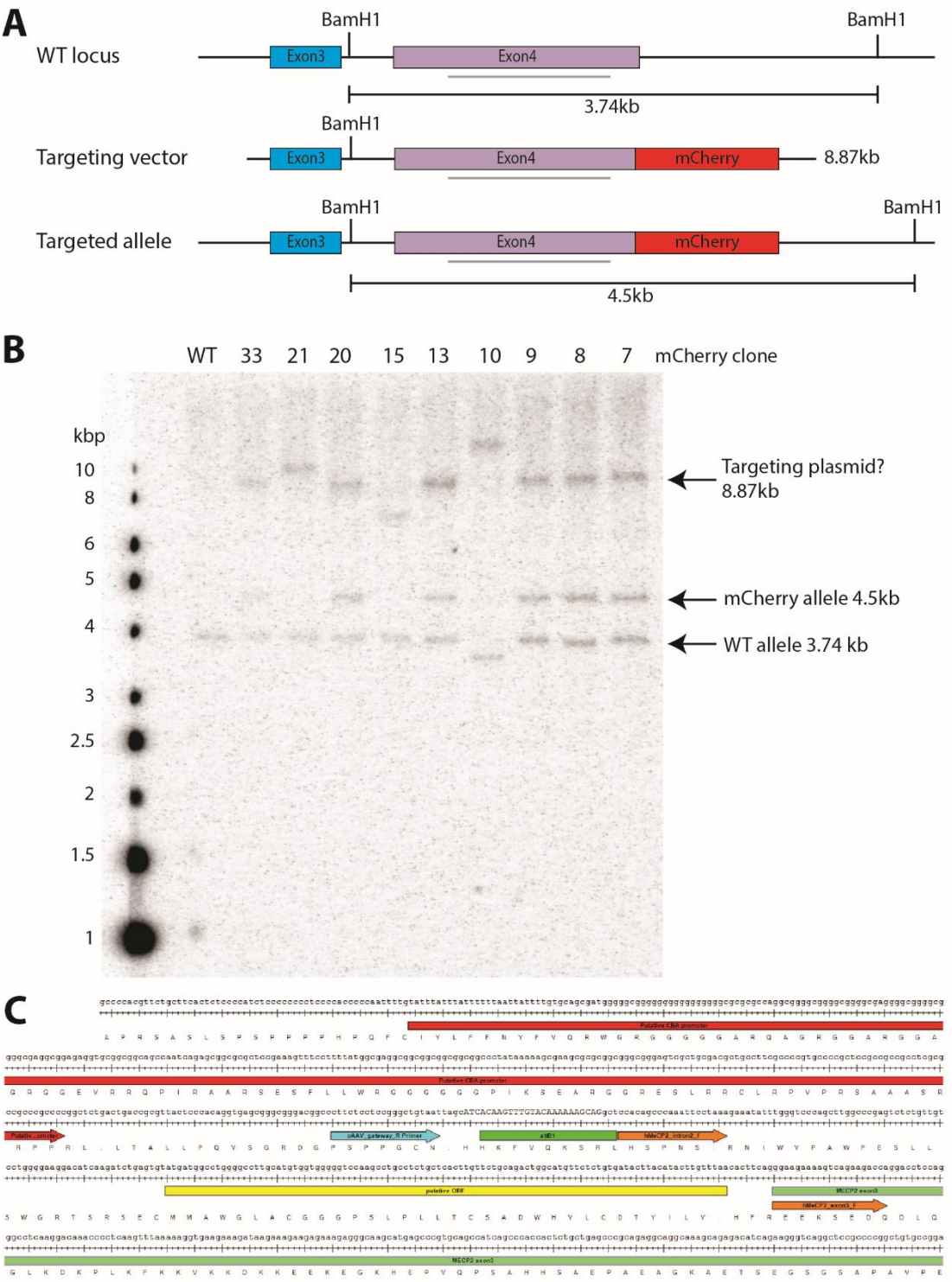
(A) hMeCP2_F and hMeCP2_R PCR of WT cells and clones that have been targeted with components shown in Figure 20, in preparation for RFLP analysis. **(B, C, D)** Schematic

representation of R111G (B), R133C (C) and T158M (D) targeting experiments. Two sgRNAs are labelled in (D), each sgRNA was used in a separate targeting experiment. Sites of DSB and PAM sites are indicated for each sgRNA, with DSB to point mutation distances highlighted. Mutations that introduce a novel restriction enzyme target sequence are also labelled.



Supplementary Figure 6. MeCP2-mCherry positive cells lines as determined by immunofluorescence imaging.

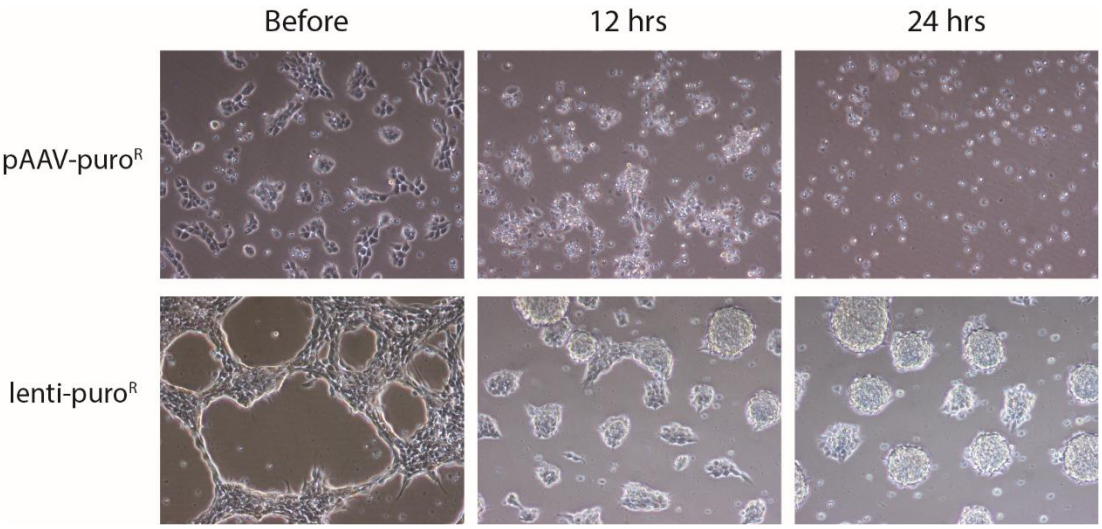
Neurons were differentiated until day 9 and then fixed and stained for neurofilament (NF), MAP2, mCherry and DAPI. Images are flattened z-stacks. Scale bar is 50 µm.



Supplementary Figure 7. Southern blot analysis of MeCP2-mCherry cell lines by Dr Jim Selfridge.

(A) Schematic of WT *MECP2* locus, targeting vector and targeted KI locus. Sites of BamH1 recognition sequences, sizes of fragments and Southern blot probe (grey) are labelled. **(B)** 0.8% gel of BamH1 Southern blot analysis of various MeCP2-mCherry cell lines (numbered according to clone number) and WT cells. Performed by Dr Jim Selfridge. **(C)** Snapshot of a portion of the mCherry targeting plasmid. The position of a putative CBA promoter is labelled in red (putative as determined by Snapgene Feature Search Tool by Dr Christina McClure), while the putative ORF

that would be encoded from this promoter is shown in yellow. This ORF begins from the first ATG that is found downstream of the promoter regardless of the reading frame and exists in intron 2 of the 5′ homology arm of the targeting vector. The 5′ homology arm begins immediately after the attB1 site (small green rectangle).

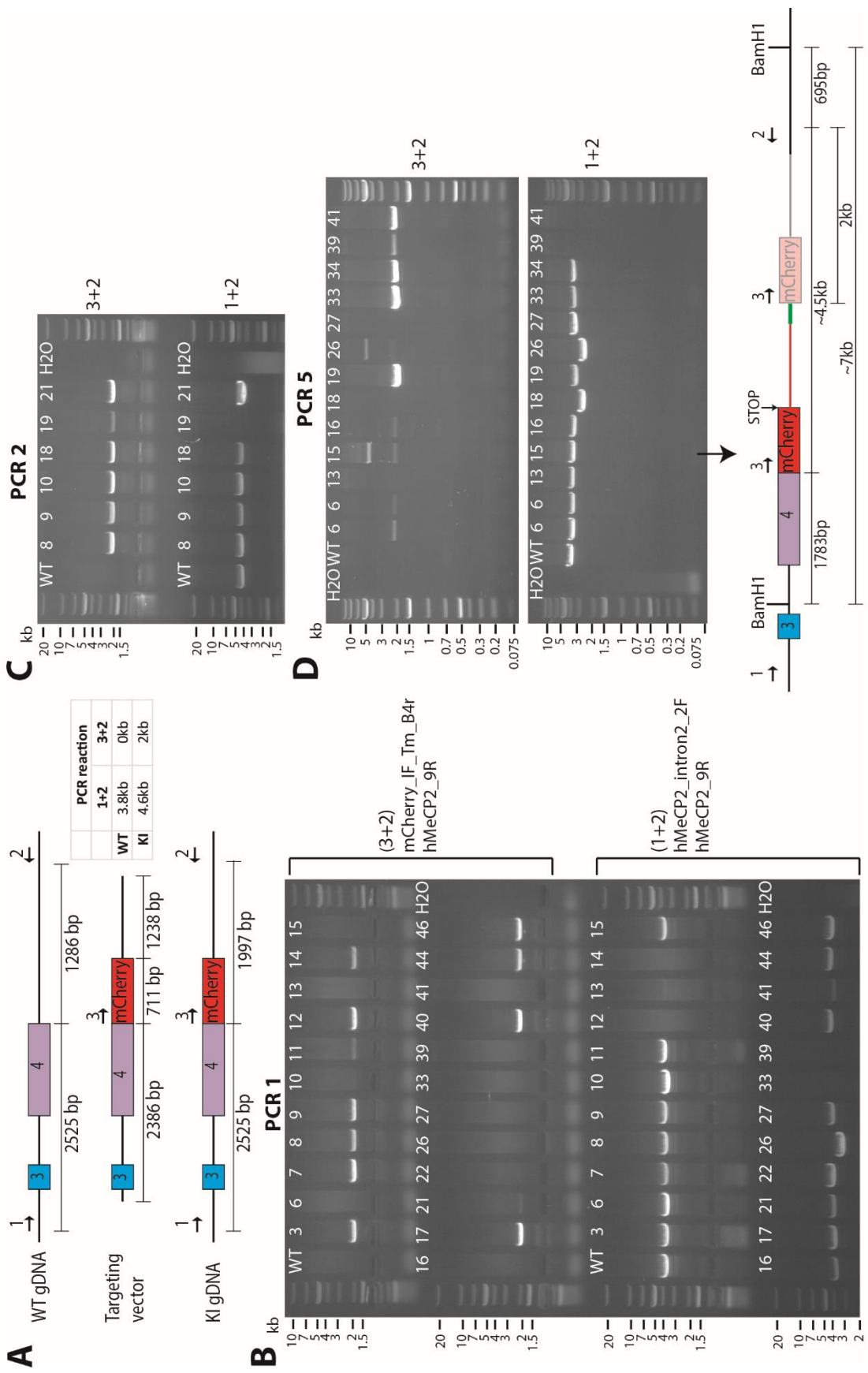


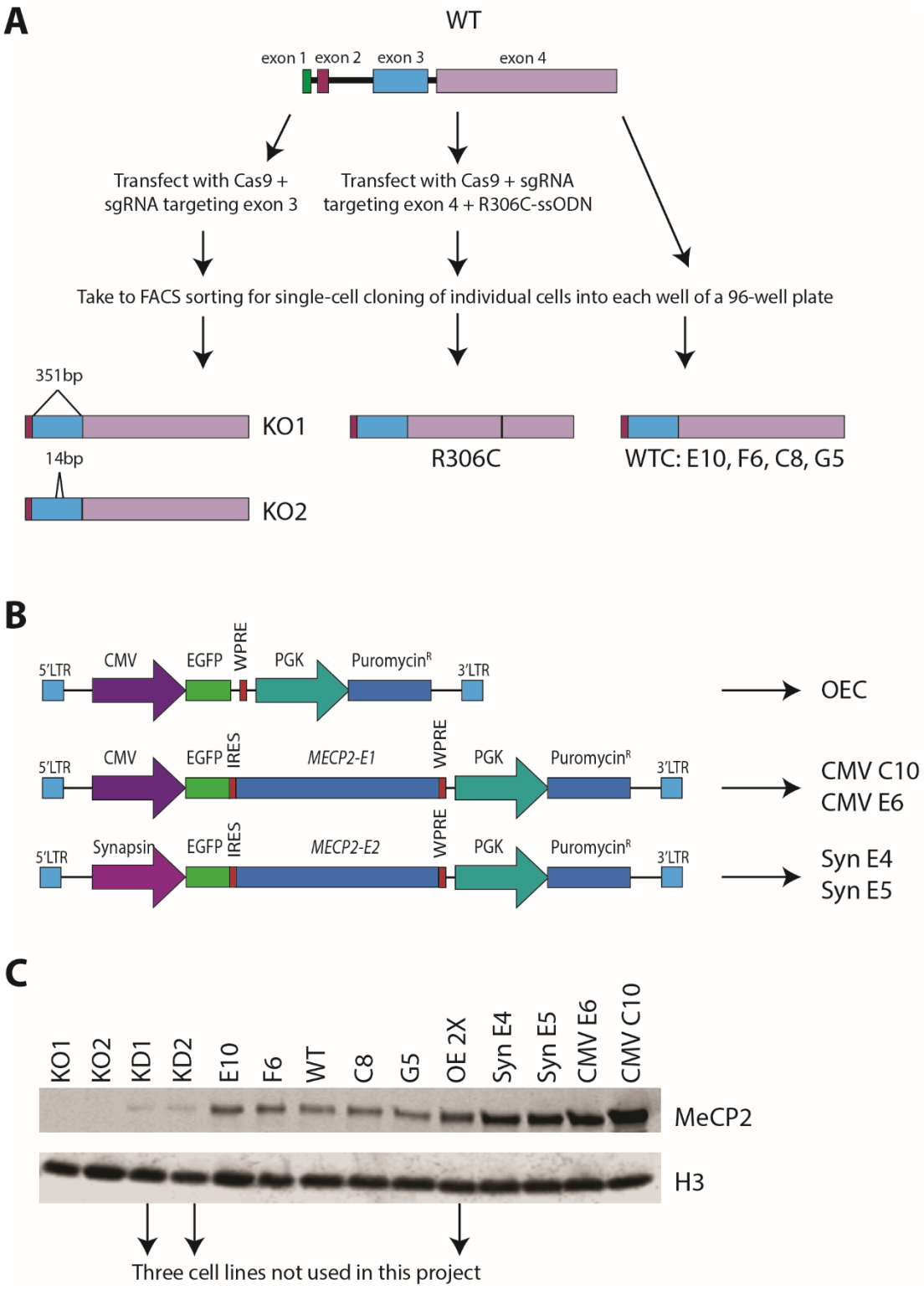
Supplementary Figure 8. Loss of puromycin resistance in cells transfected with pAAV containing a PGK-puroR-WPRE cassette.

pAAV-puroR: A pool of LUHMES cells that were transfected with the pAAV plasmid containing a PGK-puro-WPRE cassette and initially selected for based on puromycin resistance. Lenti-puroR: A clonal cell line that was infected with lentiviruses expressing PGK-puroR and are therefore constitutively puromycin resistant. Before: Condition of cell growth prior to puromycin addition. Times indicate hours after addition of 0.125 µg/ml puromycin to the cell culture medium.

Supplementary Figure 9. PCR screening of MeCP2-mCherry cell lines.

(A) Schematics of the WT locus, KI locus and the targeting vector. Locations of primers and the sgRNA site are labelled. Sizes of expected products are given. **(B)** PCR screen number 1. **(C)** PCR screen number 2. **(D)** PCR screen number 5. Schematic of potential genomic arrangement in clone 15 is shown: 3′ homology arm has been inserted twice (red and grey), with some intervening random nucleotide sequence (green; could also be inserted immediately after exon 4). PCR screen numbers correspond to Table 12. WT: genomic DNA from wild-type LUHMES cells. H_2O: water instead of DNA was used in the PCR reaction mix.

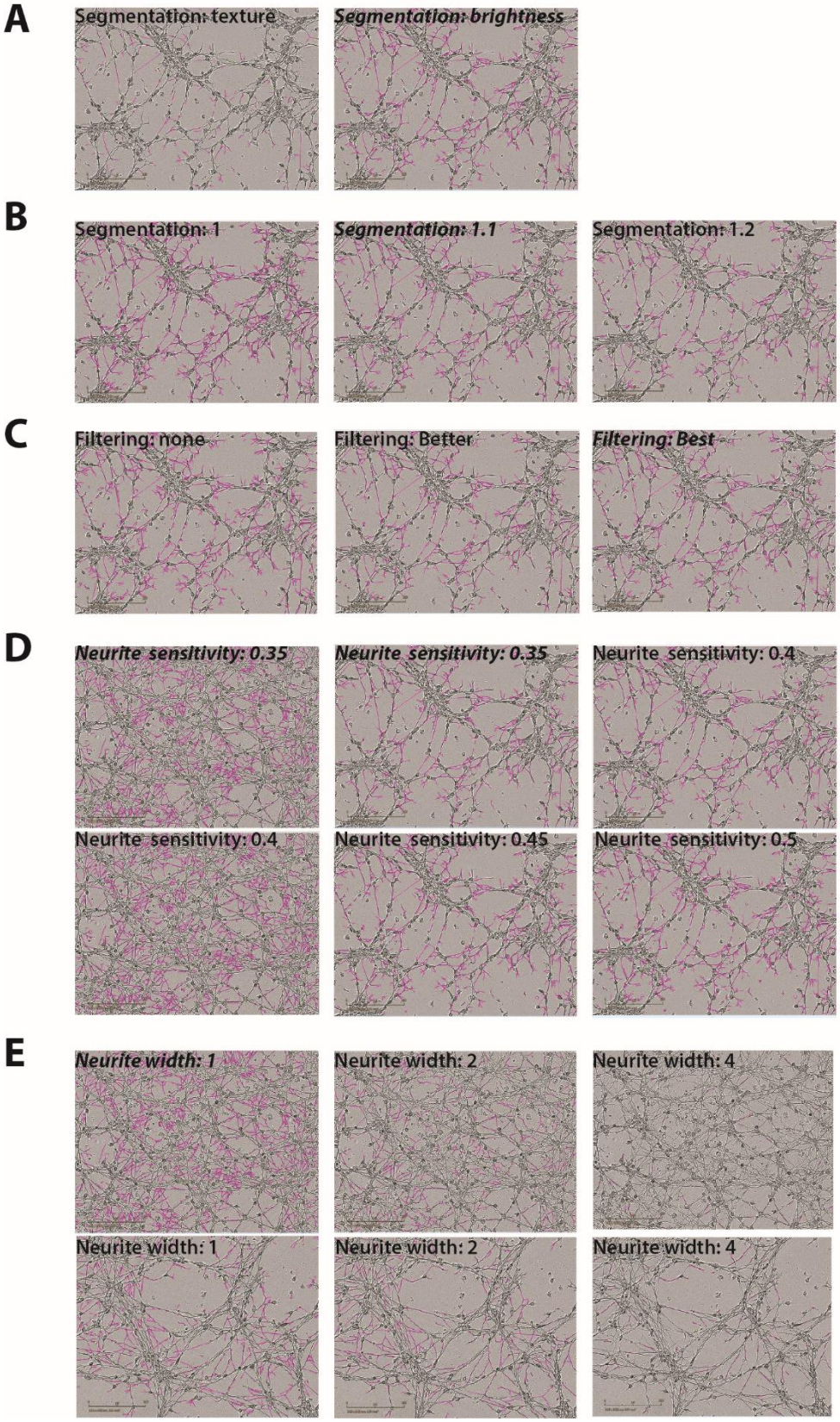

A
WT gDNA
Targeting vector
KI gDNA
2525 bp
1286 bp
2386 bp
711 bp
1238 bp
2525 bp
1997 bp
PCR reaction
1+2
3+2
WT 3.8kb 0kb
KI 4.6kb 2kb
B
PCR 1
(3+2) mCherry_IF_Tm_B4r hMeCP2_9R
(1+2) hMeCP2_intron2_2F hMeCP2_9R
C
PCR 2
3+2
1+2
D
PCR 5
3+2
1+2
BamH1
STOP
mCherry
1783bp
~4.5kb
695bp
2kb
~7kb



Supplementary Figure 10. Schematics of MeCP2 mutant cell lines.

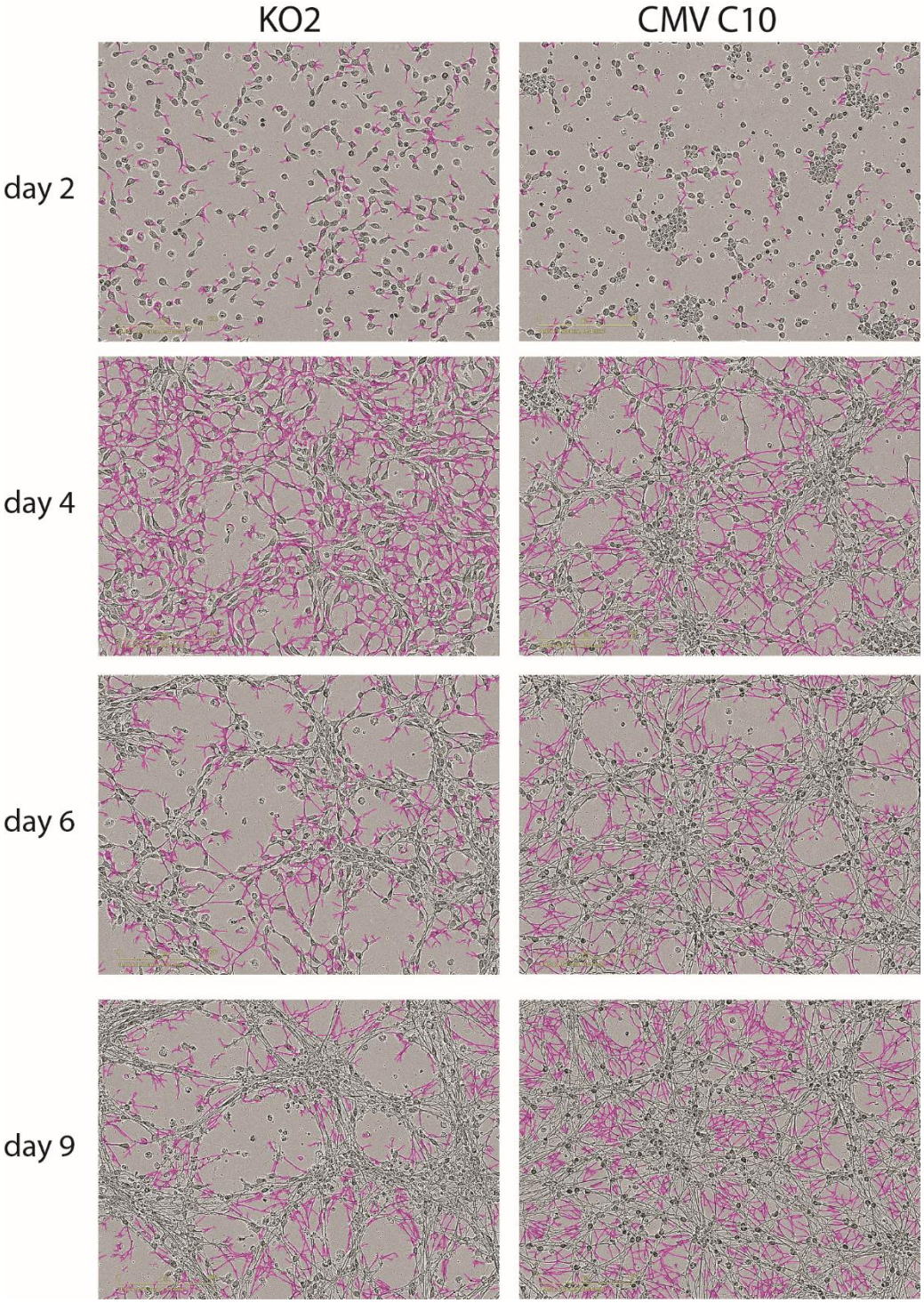
(A) Schematic and flow diagram showing generation of clonal cell lines used in this study. The full *MECP2* gene is shown for the WT schematic. *MECP2* mRNA is shown in all other schematics. **(B)** Schematics of the lentiviruses used by Dr Justyna Cholewa-Waclaw to create MeCP2 overexpression cell lines and a suitable control. **(C)** Western blot of MeCP2 KO cell lines, control cell

lines, and MeCP2 overexpressing cell lines. KD1, KD2, OE 2X are three cell lines that are not described in this project. MeCP2 antibody is M7443 (Sigma), H3 is probed for a loading control.



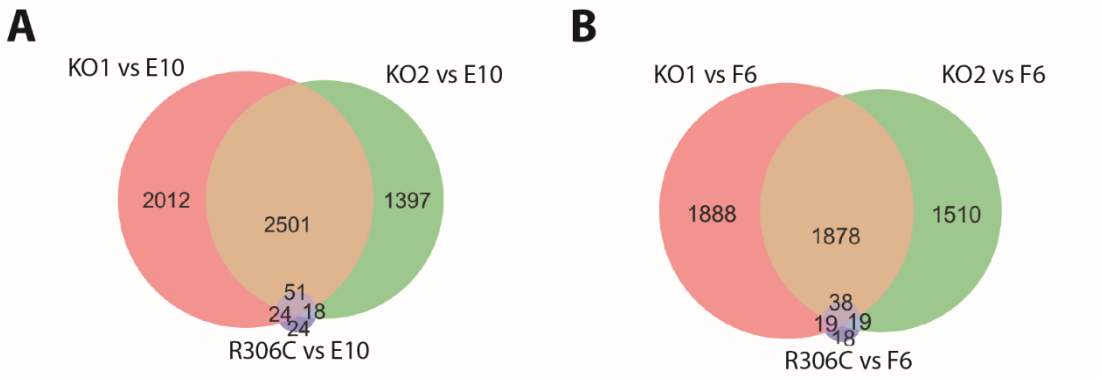
Supplementary Figure 11. Assessment of NeuroTrack software parameters to accurately define neurites in LUHMES-derived neuronal samples.

(A) Comparison of 2 segmentation modes. **(B)** Comparison of 3 segmentation adjustment values **(C)** Comparison of 3 neurite filtering parameters. **(D)** Comparison of 4 neurite sensitivity levels. **(E)** Comparison of 3 neurite width cut-off values. Scale bars are 200 µm. Parameters chosen for All Older processing definition are shown in italics.



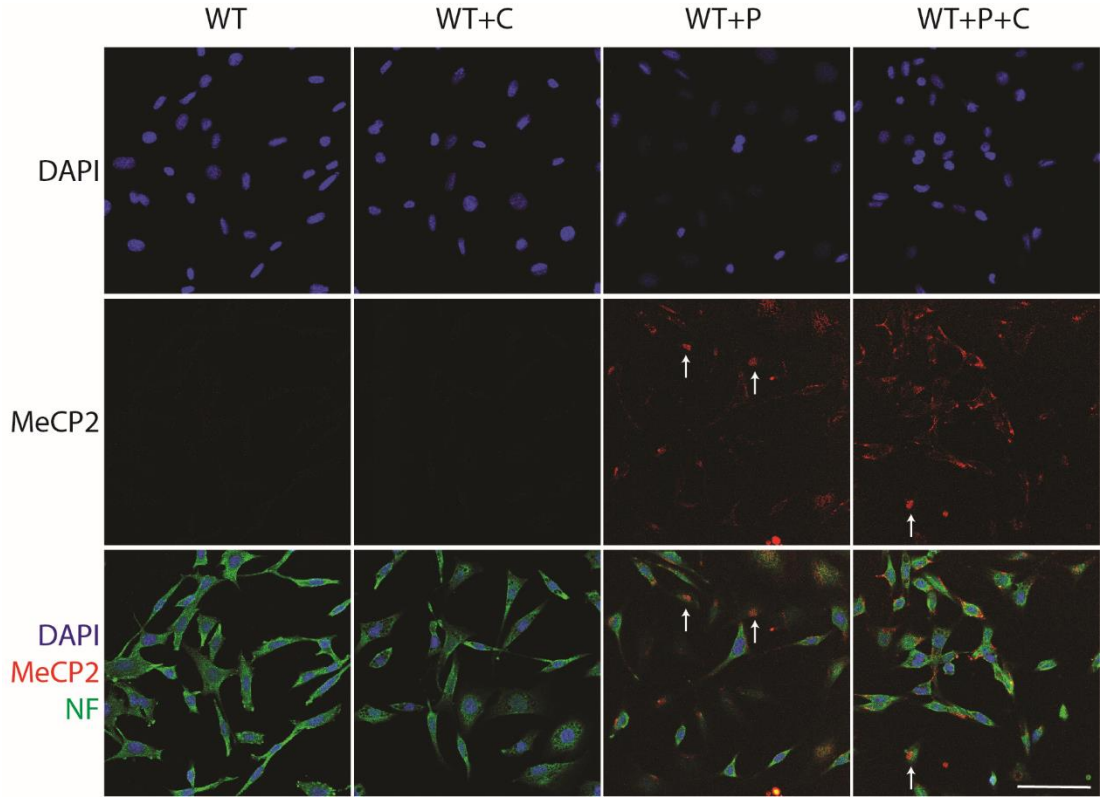
Supplementary Figure 12. Assessment of the All Older processing definition to define neurites during all stages of differentiation in two different clonal cell lines.

Phase contrast images of cell lines KO2 and CMV C10 at days 2, 6, 4 and 9 of differentiation with the neurite mask overlaid in purple. Scale bars are 200 µm.



Supplementary Figure 13. Venn diagram analysis comparing KO and R306C samples.

(A) Venn diagram comparing KO1, R306C + KO2 samples to control E10. **(B)** Venn diagram comparing KO1, R306C + KO2 samples to control F6. All significantly changing genes in each Venn diagram determined by p-adj ≤ 0.05.



Supplementary Figure 14. MeCP2 protein transduction in MEFs as judged by immunofluorescence.

(A) IF images of WT MEFs (which contain very low levels of MeCP2 protein) incubated with choloroquine (C) or MeCP2 protein (P) or both (P+C) for 4 hours, before washing, fixation and antibody staining. MeCP2 antibody is D4F3 (Cell Signalling). Scale bar is 100 µm.

References

Adams VH, McBryant SJ, Wade PA, Woodcock CL, Hansen JC (2007) Intrinsic disorder and autonomous domain function in the multifunctional nuclear protein, MeCP2. *The Journal of biological chemistry* **282:** 15057-15064

Amir RE, Veyver IBVd, Wan M, Tran CQ, Francke U, Zoghbi HY (1999) Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nature genetics* **23:** 185 - 188

Ananiev G, Williams EC, Li H, Chang Q (2011) Isogenic pairs of wild type and mutant induced pluripotent stem cell (iPSC) lines from Rett syndrome patients as in vitro disease model. *PloS one* **6:** e25255

Appanah R, Dickerson DR, Goyal P, Groudine M, Lorincz MC (2005) An Unmethylated 3' Promoter-proximal Region is Required for Efficient Transcription Initiation. *PLoS Genetics* **preprint:** e27

Ardhanareeswaran K, Mariani J, Coppola G, Abyzov A, Vaccarino FM (2017) Human induced pluripotent stem cells for modelling neurodevelopmental disorders. *Nature reviews Neurology* **13:** 265-278

Armstrong DD (2002) Neuropathology of Rett syndrome. *Mental retardation and developmental disabilities research reviews* **8:** 72-76

Assur Z, Hendrickson WA, Mancia F (2012) Tools for coproducing multiple proteins in mammalian cells. *Methods in molecular biology* **801:** 173-187

Baker SA, Chen L, Wilkins AD, Yu P, Lichtarge O, Zoghbi HY (2013) An AT-hook domain in MeCP2 determines the clinical course of Rett syndrome and related disorders. *Cell* **152:** 984-996

Ball MP, Li JB, Gao Y, Lee JH, LeProust EM, Park IH, Xie B, Daley GQ, Church GM (2009) Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nature biotechnology* **27:** 361-368

Ballestar E, Ropero S, Alaminos M, Armstrong J, Setien F, Agrelo R, Fraga MF, Herranz M, Avila S, Pineda M, Monros E, Esteller M (2005) The impact of MECP2 mutations in the expression patterns of Rett syndrome patients. *Hum Genet* **116:** 91-104

Balmer D, Goldstine J, Rao YM, LaSalle JM (2003) Elevated methyl-CpG-binding protein 2 expression is acquired during postnatal human brain development and is correlated with alternative polyadenylation. *Journal of molecular medicine* **81:** 61-68

Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P (2007) CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science* **315:** 1709-1712

Baubec T, Colombo DF, Wirbelauer C, Schmidt J, Burger L, Krebs AR, Akalin A, Schubeler D (2015) Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature*

Baubec T, Ivanek R, Lienert F, Schubeler D (2013) Methylation-dependent and -independent genomic targeting principles of the MBD protein family. *Cell* **153:** 480-492

Bebbington A, Anderson A, Ravine D, Fyfe S, Pineda M, de Klerk N, Ben-Zeev B, Yatawara N, Percy A, Kaufmann WE, Leonard H (2008) Investigating genotype-phenotype relationships in Rett syndrome using an international data set. *Neurology* **70:** 868-875

Belichenko PV, Wright EE, Belichenko NP, Masliah E, Li HH, Mobley WC, Francke U (2009) Widespread changes in dendritic and axonal morphology in Mecp2-mutant mouse models of Rett syndrome: evidence for disruption of neuronal networks. *The Journal of comparative neurology* **514:** 240-258

Ben-Hattar J, Beard P, Jiricny J (1989) Cytosine methylation in CTF and Spl recognition sites of an HSV tk promoter: effects on transcription in vivo and on factor binding in vitro. *NAR* **17:** 10179-10190

Ben-Shachar S, Chahrour M, Thaller C, Shaw CA, Zoghbi HY (2009) Mouse models of MeCP2 disorders share gene expression changes in the cerebellum and hypothalamus. *Human molecular genetics* **18:** 2431-2442

Bennett J, Ashtari M, Wellman J, Marshall KA, Cyckowski LL, Chung DC, McCague S, Pierce EA, Chen Y, Bennicelli JL, Zhu X, Ying GS, Sun J, Wright JF, Auricchio A, Simonelli F, Shindler KS, Mingozzi F, High KA, Maguire AM (2012) AAV2 gene therapy readministration in three adults with congenital blindness. *Science translational medicine* **4:** 120ra115

Bird A, Taggart M, Frommer M, Miller OJ, MacLeod D (1985) A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell* **40:** 91-99

Bird A, Tate P, Nan X, Campoy J, Meehan R, Cross S, Tweedie S, Charlton J, Macleod D (1995) Studies of DNA methylation in mammals. *Journal of cell science***:** 37-39

Bird AP (1978) Use of restriction enzymes to study eukaryotic DNA methylation. *Journal of Molecular Biology* **118:** 49-60

Bird AP (1986) CpG-rich islands and the function of DNA methylation. *Nature* **321**

Bird AP, Taggert MH (1980) Variable patterns of total DNA and rDNA methyation patterns in mammals. *NAR* **8:** 1485-1497

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30:** 2114-2120

Bolotin A, Quinquis B, Sorokin A, Ehrlich SD (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151:** 2551-2561

Boyes J, Bird A (1991) DNA Methylation Inhibits Transcription Indirectly via a Methyl-CpG Binding Protein. *Cell* **64:** 1123-1134

Brendel C, Belakhov V, Werner H, Wegener E, Gartner J, Nudelman I, Baasov T, Huppke P (2011) Readthrough of nonsense mutations in Rett syndrome: evaluation of novel aminoglycosides and generation of a new mouse model. *Journal of molecular medicine* **89:** 389-398

Broccoli V, Rubio A, Taverna S, Yekhlef L (2015) Overcoming the hurdles for a reproducible generation of human functionally mature reprogrammed neurons. *Experimental Biology and Medicine* **240:** 787-798

Brown K, Selfridge J, Lagger S, Connelly J, De Sousa D, Kerr A, Webb S, Guy J, Merusi C, Koerner MV, Bird A (2016) The molecular basis of variable phenotypic severity among common missense mutations causing Rett syndrome. *Human molecular genetics* **25:** 558-570

Bryant LM, Christopher DM, Giles AR, Hinderer C, Rodriguez JL, Smith JB, Traxler EA, Tycko J, Wojno AP, Wilson JM (2013) Lessons learned from the clinical development and market authorization of Glybera. *Human gene therapy Clinical development* **24:** 55-64

Bu Q, Wang A, Hamzah H, Waldman A, Jiang K, Dong Q, Li R, Kim J, Turner D, Chang Q (2017) CREB Signaling Is Involved in Rett Syndrome Pathogenesis. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **37:** 3671-3685

Buning H, Perabo L, Coutelle O, Quadt-Humme S, Hallek M (2008) Recent developments in adeno-associated virus vector technology. *The journal of gene medicine* **10:** 717-733

Buschhausen G, Graessmann M, Graessmann A (1985) Inhibition of herpes simplex thymidine kinase gene expression by DNA methylation is an Indirect effect. *NAR* **13:** 5503-5513

Buschhausen G, Wittig B, Graessman M, Graessmann A (1987) Chromatin structure is required to block transcription of the methylated herpes simplex virus thymidine kinase gene. *PNAS* **84:** 1177-1181

Busslinger M, Hurst J, Flavell FA (1983) DNA methylation and the Regulation of Globin Gene Expression. *Cell* **34:** 197-206

Caballero IMn, Hansen J, Leaford D, Pollard S, Hendrich BD (2009) The Methyl-CpG Binding Proteins Mecp2, Mbd2 and Kaiso Are Dispensable for Mouse Embryogenesis, but Play a Redundant Function in Neural Differentiation. *PloS one*

Calias P, Banks WA, Begley D, Scarpa M, Dickson P (2014) Intrathecal delivery of protein therapeutics to the brain: a critical reassessment. *Pharmacology & therapeutics* **144:** 114-122

Cao L, McDonnell A, Nitzsche A, Alexandrou A, Saintot P-P, Loucif AJC, Brown AR, Young G, Mis M, Randall A, Waxman SG, Stanley P, Kirby S, Tarabar S, Gutteridge A, Butt R, McKernan RM, Whiting P, Ali Z, Bilsland J, Stevens EB (2016) Pharmacological reversal of a pain phenotype in iPSC-derived sensory neurons and patients with inherited erythromelalgia. *Pain*

Chahrour M, Jung SY, Shaw C, Zhou X, Wong ST, Qin J, Zoghbi HY (2008) MeCP2, a key contributor to neurological disease, activates and represses transcription. *Science* **320:** 1224-1229

Chaligne R, Popova T, Mendoza-Parra MA, Saleem MA, Gentien D, Ban K, Piolot T, Leroy O, Mariani O, Gronemeyer H, Vincent-Salomon A, Stern MH, Heard E (2015) The inactive X chromosome is epigenetically unstable and transcriptionally labile in breast cancer. *Genome research* **25:** 488-503

Chandrasekaran A, Avci HX, Leist M, Kobolak J, Dinnyes A (2016) Astrocyte Differentiation of Human Pluripotent Stem Cells: New Tools for Neurological Disorder Research. *Frontiers in cellular neuroscience* **10:** 215

Chao HT, Chen H, Samaco RC, Xue M, Chahrour M, Yoo J, Neul JL, Gong S, Lu HC, Heintz N, Ekker M, Rubenstein JL, Noebels JL, Rosenmund C, Zoghbi HY (2010) Dysfunction in GABA signalling mediates autism-like stereotypies and Rett syndrome phenotypes. *Nature* **468:** 263-269

Charpentier E, Doudna JA (2013) Rewriting a genome. *Nature*

Chaumeil J, Augui S, Chow JC, Heard E (2008) Chapter 18: Combined Immunofluorescence, RNA Fluorescent In Situ Hybridization, and DNA Fluorescent In Situ Hybridization to Study Chromatin Changes, Transcriptional Activity, Nuclear Organization, and X-Chromosome Inactivation. *The Nucleus: Volume 1: Nuclei and Subnuclear Components*

Chen L, Chen K, Lavery LA, Baker SA, Shaw CA, Li W, Zoghbi HY (2015) MeCP2 binds to non-CG methylated DNA as neurons mature, influencing transcription and the timing of onset for Rett syndrome. *PNAS* **112:** 5509-5514

Chen RZ, Akbarian S, Tudor M, Jaenisch R (2001) Deficiency of methyl-CpG binding protein-2 in CNS neurons results in a Rett-like phenotype in mice. *Nat Genetics*

Chen WG, Chang Q, Lin Y, Meissner A, West AE, Griffith EC, Jaenisch R, Greenberg ME (2003) Derepression of BDNF transcription involves calcium-dependent phosphorylation of MeCP2. *Science* **302:** 885-889

Chen Y, Yu J, Niu Y, Qin D, Liu H, Li G, Hu Y, Wang J, Lu Y, Kang Y, Jiang Y, Wu K, Li S, Wei J, He J, Wang J, Liu X, Luo Y, Si C, Bai R, Zhang K, Liu J, Huang S, Chen Z, Wang S, Chen X, Bao X, Zhang Q, Li F, Geng R, Liang A, Shen D, Jiang T, Hu X, Ma Y, Ji W, Sun YE (2017) Modeling Rett Syndrome Using TALEN-Edited MECP2 Mutant Cynomolgus Monkeys. *Cell* **169:** 945-955 e910

Cheng TL, Chen J, Wan H, Tang B, Tian W, Liao L, Qiu Z (2017) Regulation of mRNA splicing by MeCP2 via epigenetic modifications in the brain. *Scientific reports* **7:** 42790

Cheng TL, Wang Z, Liao Q, Zhu Y, Zhou WH, Xu W, Qiu Z (2014) MeCP2 Suppresses Nuclear MicroRNA Processing and Dendritic Growth by Regulating the DGCR8/Drosha Complex. *Developmental cell* **28:** 547-560

Cheung AY, Horvath LM, Grafodatskaya D, Pasceri P, Weksberg R, Hotta A, Carrel L, Ellis J (2011) Isolation of MECP2-null Rett Syndrome patient hiPS cells and isogenic controls through X-chromosome inactivation. *Human molecular genetics* **20:** 2103-2115

Cheval H, Guy J, Merusi C, De Sousa D, Selfridge J, Bird A (2012) Postnatal inactivation reveals enhanced requirement for MeCP2 at distinct age windows. *Human molecular genetics* **21:** 3806-3814

Choi HK, Choi KC, Yoo JY, Song M, Ko SJ, Kim CH, Ahn JH, Chun KH, Yook JI, Yoon HG (2011) Reversible SUMOylation of TBL1-TBLR1 regulates beta-catenin-mediated Wnt signaling. *Mol Cell* **43:** 203-216

Choi J, Lee S, Mallard W, Clement K, Tagliazucchi GM, Lim H, Choi IY, Ferrari F, Tsankov AM, Pop R, Lee G, Rinn JL, Meissner A, Park PJ, Hochedlinger K (2015) A comparison of genetically matched cell lines reveals the equivalence of human iPSCs and ESCs. *Nature biotechnology* **33:** 1173-1181

Christodoulou J, Ho G. (2003) RettBASE: RettSyndrome.org Variation Database. Vol. 2017.

Cleary MD, Meiering CD, Jan E, Guymon R, Boothroyd JC (2005) Biosynthetic labeling of RNA with uracil phosphoribosyltransferase allows cell-specific microarray analysis of mRNA synthesis and decay. *Nature biotechnology* **23:** 232-237

Cohen S, Gabel HW, Hemberg M, Hutchinson AN, Sadacca LA, Ebert DH, Harmin DA, Greenberg RS, Verdine VK, Zhou Z, Wetsel WC, West AE, Greenberg ME (2011) Genome-wide activity-dependent MeCP2 phosphorylation regulates nervous system development and function. *Neuron* **72:** 72-85

Colantuoni C, Jeon OH, Hyder K, Chenchik A, Khimani AH, Narayanan V, Hoffman EP, Kaufmann WE, Naidu S, Pevsner J (2001) Gene expression profiling in postmortem Rett Syndrome brain: differential gene expression and patient classification. *Neurobiology of disease* **8:** 847-865

Collins AL, Levenson JM, Vilaythong AP, Richman R, Armstrong DL, Noebels JL, David Sweatt J, Zoghbi HY (2004) Mild overexpression of MeCP2 causes a progressive neurological disorder in mice. *Human molecular genetics* **13:** 2679-2689

Comb M, M.Goodman H (1990) CpG methylation inhibits proenkephalin gene expression and binding of the transcription factor AP-2. *NAR* **18:** 3975-3982

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* **339:** 819-823

Consortium IHG (2001) Initial Sequencing and Analysis of the Human Genome. *Nature* **409:** 860-921

Corbel C, Diabangouaya P, Gendrel AV, Chow JC, Heard E (2013) Unusual chromatin status and organization of the inactive X chromosome in murine trophoblast giant cells. *Development* **140:** 861-872

Core LJ, Waterfall JJ, Lis JT (2008) Nascent RNA Sequencing Reveals Widespread Pausing and Divergent Initiation at Human Promoters. *Science* **322:** 1845-1848

Coy JF, Sedlacek Z, Bächner D, Delius H, Poustka A (1999) A complex pattern of evolutionary conservation and alternative polyadenylation within the long 3'- untranslated region of themethyl-CpG-binding protein 2 gene (MECP2) suggests a regulatory role in gene expression. *Human molecular genetics* **8:** 1253-1262

Cross SH, Meehan RR, Nan X, Bird A (1997) A component of the transcriptional repressor MeCP1 shared a motif with DNA methyltransferase and HRX proteins. *Nature genetics* **16:** 256-259

Cuddapah VA, Pillai RB, Shekar KV, Lane JB, Motil KJ, Skinner SA, Tarquinio DC, Glaze DG, McGwin G, Kaufmann WE, Percy AK, Neul JL, Olsen ML (2014) Methyl-CpG-binding protein 2 (MECP2) mutation type is associated with disease severity in Rett syndrome. *Journal of medical genetics* **51:** 152-158

Curtis ARJ, Headland S, Lindsay S, Thomas NST, Boye E, Kamakari S, Roustan P, Anvret M, Wahlstrom J, McCarthy G, Clarke AJ, Bbattacharya S (1993) X chromosome linkage studies in familial Rett syndrome. *Human Genetics* **90:** 551-555

D'Astolfo DS, Pagliero RJ, Pras A, Karthaus WR, Clevers H, Prasad V, Lebbink RJ, Rehmann H, Geijsen N (2015) Efficient intracellular delivery of native proteins. *Cell* **161:** 674-690

D'Esposito M, Quaderi NA, Ciccodicola A, Bruni P, Esposito T, D'Urso M, Brown SDM (1996) Isolation, physical mapping, and Northern analysis of the X-linked human gene encoding methyl CpG-binding protein, MECP2. *Mammalian Genome* **7:** 533-535

Damian M, Porteus MH (2013) A crisper look at genome editing: RNA-guided genome modification. *Molecular therapy : the journal of the American Society of Gene Therapy* **21:** 720-722

Dawna Armstrong M (2005) Neuropathology of Rett syndrome. *Journal of Child Neurology* **20:** 747-753

Dawna Armstrong M, J Kay Dunn P, Barbara Antalffy AIMLT, Renuka Triveldi P (1995) Selective Dendritic Alterations in the cortex of Rett syndrome *Journal of Neuropathology and Developmental Biology* **54:** 195-201

Daya S, Berns KI (2008) Gene therapy using adeno-associated virus vectors. *Clinical microbiology reviews* **21:** 583-593

Delgado IJ, Kim DS, Thatcher KN, LaSalle JM, Van den Veyver IB (2006) Expression profiling of clonal lymphocyte cell cultures from Rett syndrome patients. *BMC medical genetics* **7:** 61

Deng V, Matagne V, Banine F, Frerking M, Ohliger P, Budden S, Pevsner J, Dissen GA, Sherman LS, Ojeda SR (2007) FXYD1 is an MeCP2 target gene overexpressed in the brains of Rett syndrome patients and Mecp2-null mice. *Human molecular genetics* **16:** 640-650

Derossi D, Joliot MH, Chassaing G, Prochian M (1994) The Third Helix of the Antennapedia Homeodornain Translocates through Biological ~embrane. *The Journal of biological chemistry* **269:** 10444-10450

Deyle DR, Russell DW (2009) Adeno-associated virus vector integration. *Curr Opin Mol Ther* **11:** 442-447

Di Virgilio F, Steinberg TH, Silverstein SC (1990) Inhibition of Fura-2 sequestration and secretion with organic anion transport blockers. *Cell Calcium* **11:** 57-62

Ding Q, Regan SN, Xia Y, Oostrom LA, Cowan CA, Musunuru K (2013) Enhanced efficiency of human pluripotent stem cell genome editing through replacing TALENs with CRISPRs. *Cell stem cell* **12:** 393-394

Dinh ND, Chiang YY, Hardelauf H, Baumann J, Jackson E, Waide S, Sisnaiske J, Frimat JP, van Thriel C, Janasek D, Peyrin JM, West J (2013) Microfluidic construction of minimalistic neuronal co-cultures. *Lab on a chip* **13:** 1402-1412

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29:** 15-21

Dolmetsch R, Geschwind DH (2011) The human brain in a dish: the promise of iPSC-derived neurons. *Cell* **145:** 831-834

Dragich JM, Kim YH, Arnold AP, Schanen NC (2007) Differential distribution of the MeCP2 splice variants in the postnatal mouse brain. *The Journal of comparative neurology* **501:** 526-542

Drewell RA, Goddard CJ, Thomas JO, Surani MA (2002) Methylation-dependent silencing at the imprinting control region by MeCP2. *NAR* **30:** 1139-1144

Du F, Nguyen MVC, Karten A, Felice CA, Mandel G, Ballas N (2016) Acute and crucial requirement for MeCP2 function upon transition from early to late adult stages of brain maturation. *Human molecular genetics*

Dunn HG, Stoessl AJ, Ho HH, MacLeod PM, Poskitt KJ, Doudet DJ, Schulzer M, Blackstock D, Dobko T, Koop B, Amorim GVd (2014) Rett Syndrome: Investigation of Nine Patients, including PET Scan. *The Canadian Journal of Neurological Sciences* **29:** 345-357

Ebert DH, Gabel HW, Robinson ND, Kastan NR, Hu LS, Cohen S, Navarro AJ, Lyst MJ, Ekiert R, Bird AP, Greenberg ME (2013) Activity-dependent phosphorylation of MeCP2 threonine 308 regulates interaction with NCoR. *Nature* **499:** 341-345

El-Osta A, Kantharidis P, Zalcberg JR, Wolffe AP (2002) Precipitous Release of Methyl-CpG Binding Protein 2 and Histone Deacetylase 1 from the Methylated Human Multidrug Resistance Gene (MDR1) on Activation. *Molecular and cellular biology* **22:** 1844-1857

Ellaway C, Williams K, Leonard H, Higgins G, Wilcken B, Christodoulou J (1999) Rett Syndrome: Randomized Controlled Trial of L-Carnitine. *Journal of Child Neurology* **14:** 162-167

Espuny-Camacho I, Michelsen KA, Gall D, Linaro D, Hasche A, Bonnefont J, Bali C, Orduz D, Bilheu A, Herpoel A, Lambert N, Gaspard N, Peron S, Schiffmann SN, Giugliano M, Gaillard A, Vanderhaeghen P (2013) Pyramidal neurons derived from human pluripotent stem cells integrate efficiently into mouse brain circuits in vivo. *Neuron* **77:** 440-456

Farra N, Zhang WB, Pasceri P, Eubanks JH, Salter MW, Ellis J (2012) Rett syndrome induced pluripotent stem cell-derived neurons reveal novel neurophysiological alterations. *Molecular psychiatry* **17:** 1261-1271

Feng J, Zhou Y, Campbell SL, Le T, Li E, Sweatt JD, Silva AJ, Fan G (2010) Dnmt1 and Dnmt3a maintain DNA methylation and regulate synaptic function in adult forebrain neurons. *Nature neuroscience* **13:** 423-430

Fichou Y, Nectoux J, Bahi-Buisson N, Rosas-Vargas H, Girard B, Chelly J, Bienvenu T (2009) The first missense mutation causing Rett syndrome specifically affecting the MeCP2_e1 isoform. *Neurogenetics* **10:** 127-133

Filion GJ, Zhenilo S, Salozhin S, Yamada D, Prokhortchouk E, Defossez PA (2006) A family of human zinc finger proteins that bind methylated DNA and repress transcription. *Molecular and cellular biology* **26:** 169-181

Flanagan JM, Wild L (2007) An epigenetic role for noncoding RNAs and intragenic DNA methylation. *Genome biology* **8:** 307

Forlani G, Giarda E, Ala U, Di Cunto F, Salani M, Tupler R, Kilstrup-Nielsen C, Landsberger N (2010) The MeCP2/YY1 interaction regulates ANT1 expression at 4q35: novel hints for Rett syndrome pathogenesis. *Human molecular genetics* **19:** 3114-3123

Free A, Wakefield RI, Smith BO, Dryden DT, Barlow PN, Bird AP (2001) DNA recognition by the methyl-CpG binding domain of MeCP2. *The Journal of biological chemistry* **276:** 3353-3360

Friedland AE, Tzur YB, Esvelt KM, Colaiacovo MP, Church GM, Calarco JA (2013) Heritable genome editing in C. elegans via a CRISPR-Cas9 system. *Nature methods* **10:** 741-743

Fuchs G, Voichek Y, Benjamin S, Gilad S, Amit I, Oren M (2016) 4sUDRB-seq: measuring genomewide transcriptional elongation rates and initiation frequencies within cells. *Genome biology* **15:** 1-10

Fujita N, Takebayashi S-I, Okumura K, Kudo S, Chiba T, Saya H, Nakao M (1999) Methylation-Mediated Transcriptional Silencing in Euchromatin by Methyl-CpG Binding Protein MBD1 Isoforms. *Molecular and cellular biology* **19:** 6415-6426

Futaki S, Suzuki T, Ohashi W, Yagami T, Tanaka S, Ueda K, Sugiura Y (2001) Arginine-rich peptides. An abundant source of membrane-permeable peptides having potential as carriers for intracellular protein delivery. *The Journal of biological chemistry* **276:** 5836-5840

Fyffe SL, Neul JL, Samaco RC, Chao HT, Ben-Shachar S, Moretti P, McGill BE, Goulding EH, Sullivan E, Tecott LH, Zoghbi HY (2008) Deletion of Mecp2 in Sim1-expressing neurons reveals a critical role for MeCP2 in feeding behavior, aggression, and the response to stress. *Neuron* **59:** 947-958

Gabel HW, Kinde B, Stroud H, Gilbert CS, Harmin DA, Kastan NR, Hemberg M, Ebert DH, Greenberg ME (2015) Disruption of DNA-methylation-dependent long gene repression in Rett syndrome. *Nature* **522:** 89-93

Gabriele V. Ronnett M, PhD, Linda Moses B, Donald Leopold M, Eric P. Hoffman P, Xiaohe Cai M, Kristen C. Hoffbuhr P, SakkuBai Naidu M (2003) Olfactory Biopsies Demonstrate a Defect in Neuronal Development in Rett's Syndrome. *Annals of Neurology* **54:** 206-218

Gadalla KK, Bailey ME, Spike RC, Ross PD, Woodard KT, Kalburgi SN, Bachaboina L, Deng JV, West AE, Samulski RJ, Gray SJ, Cobb SR (2013) Improved survival and reduced phenotypic severity following AAV9/MECP2 gene transfer to neonatal and juvenile male Mecp2 knockout mice. *Molecular therapy : the journal of the American Society of Gene Therapy* **21:** 18-30

Gadalla KKE, Vudhironarit T, Hector RD, Sinnett S, Bahey NG, Bailey MES, Gray SJ, Cobb SR (2017) Development of a Novel AAV Gene Therapy Cassette with Improved Safety Features and Efficacy in a Mouse Model of Rett Syndrome. *Molecular therapy Methods & clinical development* **5:** 180-190

Gaj T, Staahl BT, Rodrigues GM, Limsirichai P, Ekman FK, Doudna JA, Schaffer DV (2017) Targeted gene knock-in by homology-directed genome editing using Cas9 ribonucleoprotein and AAV donor delivery. *Nucleic acids research*

Galluzzi L, Aaronson SA, Abrams J, Alnemri ES, Andrews DW, Baehrecke EH, Bazan NG, Blagosklonny MV, Blomgren K, Borner C, Bredesen DE, Brenner C, Castedo M, Cidlowski JA, Ciechanover A, Cohen GM, De Laurenzi V, De Maria R, Deshmukh M, Dynlacht BD, El-Deiry WS, Flavell RA, Fulda S, Garrido C, Golstein P, Gougeon ML, Green DR, Gronemeyer H, Hajnoczky G, Hardwick JM, Hengartner MO, Ichijo H, Jaattela M, Kepp O, Kimchi A, Klionsky DJ, Knight RA, Kornbluth S, Kumar S, Levine B, Lipton SA, Lugli E, Madeo F, Malomi W, Marine JC, Martin SJ, Medema JP, Mehlen P, Melino G, Moll UM, Morselli E, Nagata S, Nicholson DW, Nicotera P, Nunez G, Oren M, Penninger J, Pervaiz S, Peter ME, Piacentini M, Prehn JH, Puthalakath H, Rabinovich GA, Rizzuto R, Rodrigues CM, Rubinsztein DC, Rudel T,

Scorrano L, Simon HU, Steller H, Tschopp J, Tsujimoto Y, Vandenabeele P, Vitale I, Vousden KH, Youle RJ, Yuan J, Zhivotovsky B, Kroemer G (2009) Guidelines for the use and interpretation of assays for monitoring cell death in higher eukaryotes. *Cell death and differentiation* **16:** 1093-1107

Garg SK, Lioy DT, Cheval H, McGann JC, Bissonnette JM, Murtha MJ, Foust KD, Kaspar BK, Bird A, Mandel G (2013) Systemic delivery of MeCP2 rescues behavioral and cellular deficits in female mouse models of Rett syndrome. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **33:** 13612-13620

Gaspard N, Bouschet T, Herpoel A, Naeije G, van den Ameele J, Vanderhaeghen P (2009) Generation of cortical neurons from mouse embryonic stem cells. *Nature protocols* **4:** 1454-1463

Gaudet D, Stroes ES, Méthot J, Brisson D, Tremblay K, Moens SJB, Iotti G, Rastelletti I, Ardigo D, Corzo D, Meyer C, Andersen M, Ruszniewski P, Deakin M, Bruno MJ (2016) Long-Term Retrospective Analysis of Gene Therapy with Alipogene Tiparvovec and Its Effect on Lipoprotein Lipase Deficiency-Induced Pancreatitis. *Human gene therapy* **27:** 916-925

Gendrel AV, Attia M, Chen CJ, Diabangouaya P, Servant N, Barillot E, Heard E (2014) Developmental dynamics and disease potential of random monoallelic gene expression. *Developmental cell* **28:** 366-380

Georgel PT, Horowitz-Scherer RA, Adkins N, Woodcock CL, Wade PA, Hansen JC (2003) Chromatin compaction by human MeCP2. Assembly of novel secondary chromatin structures in the absence of DNA methylation. *The Journal of biological chemistry* **278:** 32181-32188

Ghosh RP, Horowitz-Scherer RA, Nikitina T, Shlyakhtenko LS, Woodcock CL (2010) MeCP2 binds cooperatively to its substrate and competes with histone H1 for chromatin binding sites. *Molecular and cellular biology* **30:** 4656-4670

Goffin D, Allen M, Zhang L, Amorim M, Wang IT, Reyes AR, Mercado-Berton A, Ong C, Cohen S, Hu L, Blendy JA, Carlson GC, Siegel SJ, Greenberg ME, Zhou Z (2012) Rett syndrome mutation MeCP2 T158A disrupts DNA binding, protein stability and ERP responses. *Nature neuroscience* **15:** 274-283

Gossen M, Bujard H (1992) Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *PNAS* **89:** 5547-5551

Graff J, Kim D, Dobbin MM, Tsai L-H (2011) Epigenetic Regulation of Gene Expression in Physiological and Pathological Brain Processes. *PHYSIOLOGICAL REVIEWS* **91:** 603–649,

Gratz SJ, Cummings AM, Nguyen JN, Hamm DC, Donohue LK, Harrison MM, Wildonger J, O'Connor-Giles KM (2013) Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease. *Genetics*

Guo JU, Su Y, Shin JH, Shin J, Li H, Xie B, Zhong C, Hu S, Le T, Fan G, Zhu H, Chang Q, Gao Y, Ming GL, Song H (2013) Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. *Nature neuroscience*

Guy J, Cheval H, Selfridge J, Bird A (2011) The role of MeCP2 in the brain. *Annual review of cell and developmental biology* **27:** 631-652

Guy J, Gan J, Selfridge J, Cobb S, Bird A (2007) Reversal of neurological defects in a mouse model of Rett syndrome. *Science* **315:** 1143-1147

Guy J, Hendrich B, Holmes M, Martin JE, Bird A (2001) A mouse Mecp2-null mutation causes neurological symptoms that mimic Rett syndrome. *Nature genetics* **27:** 322-326

Hagberg B, Aicardi J, Dias K, Ramos O (1983) A Progressive Syndrome of Autism, Dementia, Ataxia, and Loss of Purposful Hand Use in Girls: Rett's Syndrome: Report of 35 Cases. *Ann Neurol* **14:** 471-479

Hai T, Teng F, Guo R, Li W, Zhou Q (2014) One-step generation of knockout pigs by zygote injection of CRISPR/Cas system. *Cell research*

Halbert CL, Standaert TA, Wilson CB, Miller AD (1998) Successful Readministration of Adeno-Associated Virus Vectors to the Mouse Lung Requires Transient Immunosuppression during the Initial Exposure. *Journal of virology* **72:** 9795-9805

Hansen JC, Ghosh RP, Woodcock CL (2010) Binding of the Rett syndrome protein, MeCP2, to methylated and unmethylated DNA and chromatin. *IUBMB life* **62:** 732-738

Harikrishnan KN, Bayles R, Ciccotosto GD, Maxwell S, Cappai R, Pelka GJ, Tam PP, Christodoulou J, El-Osta A (2010) Alleviating transcriptional inhibition of the norepinephrine slc6a2 transporter gene in depolarized neurons. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **30:** 1494-1501

Harikrishnan KN, Chow MZ, Baker EK, Pal S, Bassal S, Brasacchio D, Wang L, Craig JM, Jones PL, Sif S, El-Osta A (2005) Brahma links the SWI/SNF chromatin-remodeling complex with MeCP2-dependent transcriptional silencing. *Nature genetics* **37:** 254-264

Heckman LD, Chahrour MH, Zoghbi HY (2014) Rett-causing mutations reveal two domains critical for MeCP2 function and for toxicity in MECP2 duplication syndrome mice. *eLIFE*

Heidemann M, Hintermair C, Voss K, Eick D (2013) Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription. *Biochimica et biophysica acta* **1829:** 55-62

Heinen CA, Jongejan A, Watson PJ, Redeker B, Boelen A, Boudzovitch-Surovtseva O, Forzano F, Hordijk R, Kelley R, Olney AH, Pierpont ME, Schaefer GB, Stewart F, Trotsenburg ASPv, Fliers E, Schwabe JWR, Hennekam RC (2016) A specific mutation in TBL1XR1 causes Pierpont syndrome. *Journal of medical genetics*

Hendrich B, Bird A (1998) Identification and Characterization of a Family of Mammalian Methyl-CpG Binding Proteins. *Molecular and cellular biology* **18:** 6538-6547

Hendrich B, Tweedie S (2003) The methyl-CpG binding domain and the evolving role of DNA methylation in animals. *Trends in Genetics* **19:** 269-277

Hendrie PC, Russell DW (2005) Gene targeting with viral vectors. *Molecular therapy : the journal of the American Society of Gene Therapy* **12:** 9-17

Hermanson O, Jepsen K, Rosenfeld MG (2002) N-CoR controls differentiation of neural stem cells into astrocytes. *Nature* **419:** 934-939

Heyer WD, Ehmsen KT, Liu J (2010) Regulation of homologous recombination in eukaryotes. *Annual review of genetics* **44:** 113-139

Higashimoto T, Urbinati F, Perumbeti A, Jiang G, Zarzuela A, Chang L-J, Kohn D, Malik P (2007) The woodchuck hepatitis virus post-transcriptional regulatory element reduces readthrough transcription from retroviral vectors. *Gene therapy* **14:** 1298-1304

Ho KL, McNae IW, Schmiedeberg L, Klose RJ, Bird AP, Walkinshaw MD (2008) MeCP2 binding to DNA depends upon hydration at methyl-CpG. *Mol Cell* **29:** 525-531

Holliday R, Pugh JE (1975) DNA Modification Mechanisms and Gene Activity during Development. *Science* **189:** 226-232

Horike S, Cai S, Miyano M, Cheng JF, Kohwi-Shigematsu T (2005) Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. *Nature genetics* **37:** 31-40

Hoshimaru M, Ray J, Sah DD, Gage FH (1996) Differentiation of the immortalized adult neuronal progenitor cell line HC2S2 into neurons by regulatable suppression of the v-myc oncogene. *PNAS* **96:** 1518-1523

Hotchkiss RD (1948) Quantitative separation of purines, pyrimidines and nucleosides by paper chromatography. *Journal of Biological Chemistry* **175:** 315-322

Hotta A, Cheung AYL, Farra N, Vijayaragavan K, Séguin CA, Draper JS, Pasceri P, Maksakova IA, Mager DL, Rossant J, Bhatia M, Ellis J (2009) Isolation of human iPS cells using EOS lentiviral vectors to select for pluripotency. *Nature methods* **6:** 370-376

Houseley J, Tollervey D (2009) The many pathways of RNA degradation. *Cell* **136:** 763-776

Howden SE, McColl B, Glaser A, Vadolas J, Petrou S, Little MH, Elefanty AG, Stanley EG (2016) A Cas9 Variant for Efficient Generation of Indel-Free Knockin or Gene-Corrected Human Pluripotent Stem Cells. *Stem cell reports*

Hu K, Nan X, Bird A, Wang W (2006) Testing for association between MeCP2 and the brahma-associated SWI/SNF chromatin-remodeling complex. *Nature genetics* **38:** 962-964

Huang L-H, Wang R, Gama-Sosa MA, Shenoy S, Ehrlich M (1984) A protein from human placental nuclei binds preferentially to 5-methylcytosine-rich DNA. *Nature* **308:** 293-295

Hubrich-Kuhner K, Buhk H-J, Wagner H, Krogner H, Simon D (1989) Non-C-G recognition sequences of DNA cytosin-5-methyltransferase from rat liver. *Biochemical and biophysical research communications* **160:** 1175-1182

Hughes MA, Brennan PM, Bunting AS, Cameron K, Murray AF, Shipston MJ (2014) Patterning human neuronal networks on photolithographically engineered silicon dioxide substrates functionalized with glial analogues. *Journal of biomedical materials research Part A* **102:** 1350-1360

Huh CJ, Zhang B, Victor MB, Dahiya S, Batista LF, Horvath S, Yoo AS (2016) Maintenance of age in human neurons generated by microRNA-based neuronal conversion of fibroblasts. *eLIFE* **5:** 1-14

Humphreys P, Barrowman N (2016) The Incidence and Evolution of Parkinsonian Rigidity in Rett Syndrome: A Pilot Study. *The Canadian journal of neurological sciences Le journal canadien des sciences neurologiques* **43:** 567-573

Hunihan L, Brown J, Cacace A, Fernandes A, Weston A (2017) Generation of a clonal induced pluripotent stem cell (iPSC) line expressing the mutant MECP2 allele from a Rett Syndrome patient fibroblast line. *Stem Cell Res* **20:** 67-69

Hwang WY, Fu Y, Reyon D, Maeder ML, Tsai SQ, Sander JD, Peterson RT, Yeh JR, Joung JK (2013) Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature biotechnology* **31:** 227-229

Ichiyanagi T, Ichiyanagi K, Miyake M, Sasaki H (2013) Accumulation and loss of asymmetric non-CpG methylation during male germ-cell development. *Nucleic acids research* **41:** 738-745

Iguchi-Ariga SMM, Schaffner W (1989) CpG methylation of the cAMP-responsive enhancer/promoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation. *Genes & development* **3:** 612-619

Ilieva M, Della Vedova P, Hansen O, Dufva M (2013) Tracking neuronal marker expression inside living differentiating cells using molecular beacons. *Frontiers in cellular neuroscience* **7:** 266

Illingworth R, Kerr A, DeSousa D, Jørgensen H, Ellis P, Stalker J, Jackson D, Clee C, Plumb R, Rogers J, Humphray S, Cox T, Langford C, Bird A (2008) A Novel CpG Island Set Identifies Tissue-Specific Methylation at Developmental Gene Loci. *PloS Biology* **6:** 34-51

Illingworth RS, Gruenewald-Schneider U, Webb S, Kerr AR, James KD, Turner DJ, Smith C, Harrison DJ, Andrews R, Bird AP (2010) Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *PLoS Genet* **6:** e1001134

Ishibashi T, Thambirajah AA, Ausio J (2008) MeCP2 preferentially binds to methylated linker DNA in the absence of the terminal tail of histone H3 and independently of histone acetylation. *FEBS letters* **582:** 1157-1162

Ishino Y, Shingawa H, Makino K, Amemura M, Nakata A (1987) Nucleotide Sequence of the iap Gene, Responsible for Alkaline Phosphatase Isozyme Conversion in Escherichia coli, and Identification of the Gene Product. *Journal of Bacteriology* **169:** 5429-5433

Israel MA, Yuan SH, Bardy C, Reyna SM, Mu Y, Herrera C, Hefferan MP, Van Gorp S, Nazor KL, Boscolo FS, Carson CT, Laurent LC, Marsala M, Gage FH, Remes AM, Koo EH, Goldstein LS (2012) Probing sporadic and familial Alzheimer's disease using induced pluripotent stem cells. *Nature* **482:** 216-220

Ito-Ishida A, Ure K, Chen H, Swann JW, Zoghbi HY (2015) Loss of MeCP2 in Parvalbumin-and Somatostatin-Expressing Neurons in Mice Leads to Distinct Rett Syndrome-like Phenotypes. *Neuron* **88:** 651-658

J. Edmond Wraith M, CHB,, Anna Tylki-Szymanska M, PhD, Nathalie Guffon M, PhD, Y. Howard Lien M, PhD, Michel Tsimaratos M, PhD, Ashok Vellodi M, Dominique P. Germain M, PhD (2008) Safety and Efficacy of Enzyme Replacement Therapy with Agalsidase Beta: An International, Open-label Study in Pediatric Patients with Fabry Disease. *Journal of Pediatrics***:** 563-571

Jansen R, Gaastra W, Embden JDAv, Schouls LM (2002) Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular Microbiology* **43:** 1565–1575

Jentarra GM, Olfers SL, Rice SG, Srivastava N, Homanics GE, Blue M, Naidu S, Narayanan V (2010) Abnormalities of cell packing density and dendritic complexity in the MeCP2 A140V mouse model of Rett syndrome/X-linked mental retardation. *BMC Neuroscience* **11:** 1-15

Jepsen K, Hermanson O, Onami TM, Gleiberman AS, Lunyak V, Kurokawa R, McEvilly RJ, Kumar V, Liu F, Seto E, Hedrick SM, Mandel G, Glass CK, Rose DW, Rosenfeld MG (2000) Combinatorial Roles of the Nuclear Receptor Corepressor in Transcription and Development. *Cell* **102:** 753-763

Jiang M, Ash RT, Baker SA, Suter B, Ferguson A, Park J, Rudy J, Torsky SP, Chao HT, Zoghbi HY, Smirnakis SM (2013a) Dendritic arborization and spine dynamics are abnormal in the mouse model of MECP2 duplication syndrome. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **33:** 19518-19533

Jiang W, Zhou H, Bi H, Fromm M, Yang B, Weeks DP (2013b) Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. *Nucleic acids research* **41:** e188

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337:** 816-821

Jinek M, East A, Cheng A, Lin S, Ma E, Doudna J (2013) RNA-programmed genome editing in human cells. *eLIFE*

Johnson MA, Weick JP, Pearce RA, Zhang SC (2007) Functional neural development from human embryonic stem cells: accelerated synaptic activity via astrocyte coculture. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **27:** 3069-3077

Jones PA (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature reviews Genetics* **13:** 484-492

Jones PA, Liang G (2009) Rethinking how DNA methylation patterns are maintained. *Nature reviews Genetics* **10:** 805-811

Jones PA, Taylor SM (1980) Cellular Differentiation, Cytidine Analogues and DNA Methylation. *Cell* **20:** 85-93

Jones PL, Veenstra GJC, Wade PA, Vermaak D, Kass SU, Landsberger N, Strouboulis J, Wolffe AP (1998) Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nature genetics*

Jordan C, Li HH, Kwan HC, Francke U (2007) Cerebellar gene expression profiles of mouse models for Rett syndrome reveal novel MeCP2 targets. *BMC medical genetics* **8:** 36

Jorgensen HF, Ben-Porath I, Bird AP (2004) Mbd1 Is Recruited to both Methylated and Nonmethylated CpGs via Distinct DNA Binding Domains. *Molecular and cellular biology* **24:** 3387-3395

Jr. JJC, Miller MR, Bardee AP (1978) Animal cells reversibly permeable to small molecules. *PNAS* **75:** 351-355

Jugloff DG, Jung BP, Purushotham D, Logan R, Eubanks JH (2005) Increased dendritic complexity and axonal length in cultured mouse cortical neurons overexpressing methyl-CpG-binding protein MeCP2. *Neurobiology of disease* **19:** 18-27

Jung BP, Jugloff DG, Zhang G, Logan R, Brown S, Eubanks JH (2003) The expression of methyl CpG binding factor MeCP2 correlates with cellular differentiation in the developing rat brain and in cultured cells. *Journal of neurobiology* **55:** 86-96

Juopperi TA, Kim WR, Chiang CH, Yu H, Margolis RL, Ross CA, Ming GL, Song H (2012) Astrocytes generated from patient induced pluripotent stem cells recapitulate features of Huntington's disease patient cells. *Mol Brain* **5:** 17

Kabouridis PS (2003) Biological applications of protein transduction technology. *Trends in biotechnology* **21:** 498-503

Kaludov NK, P.Wolffe A (2000) MeCP2 driven transcriptional repression: selectivity for methylated DNA, action at a distance and contacts with the basal transcription machinery. *NAR* **28:** 1921–1928

Kang X, He W, Huang Y, Yu Q, Chen Y, Gao X, Sun X, Fan Y (2016) Introducing precise genetic modifications into human 3PN embryos by CRISPR/Cas-mediated genome editing. *Journal of assisted reproduction and genetics* **33:** 581-588

Kass SU, Goddard JP, Adams RLP (1993) Inactive Chromatin Spreads from a Focus of Methylation. *Molecular and cellular biology* **13:** 7372-7379

Katz DM, Bird A, Coenraads M, Gray SJ, Menon DU, Philpot BD, Tarquinio DC (2016) Rett Syndrome: Crossing the Threshold to Clinical Translation. *Trends in neurosciences* **39:** 100-113

Kaufmann WE, MacDonald SM, Altamura CR (2000) Dendritic Cytoskeletal Protein Expression in Mental Retardation: An Immunohistochemical Study of the Neocortex in Rett Syndrome. *Cerebral Cortex* **10:** 992-1004

Kaufmann WE, Stallworth JL, Everman DB, Skinner SA (2016) Neurobiologically-based treatments in Rett syndrome: opportunities and challenges. *Expert Opinion on Orphan Drugs* **4:** 1043-1055

Kerr B, Soto CJ, Saez M, Abrams A, Walz K, Young JI (2012) Transgenic complementation of MeCP2 deficiency: phenotypic rescue of Mecp2-null mice by isoform-specific transgenes. *European journal of human genetics : EJHG* **20:** 69-76

Khwaja OS, Ho E, Barnes KV, O'Leary HM, Pereira LM, Finkelstein Y, III CAN, Vogel-Farley V, DeGregorio G, Holm IA, Khatwa U, Kapur K, Alexander ME, Finnegan DM, Cantwell NG, Walco AC, Rappaport L, Gregas M, Fichorova RN, Shannon MW, Sur M, Kaufmann WE (2014) Safety, pharmacokinetics, and preliminary assessment of efficacy of mecasermin (recombinant human IGF-1) for the treatment of Rett syndrome. *PNAS* **111:** 4596-4601

Kim K, Doi A, Wen B, Ng K, Zhao R, Cahan P, Kim J, Aryee MJ, Ji H, Ehrlich LI, Yabuuchi A, Takeuchi A, Cunniff KC, Hongguang H, McKinney-Freeman S, Naveiras O, Yoon TJ, Irizarry RA, Jung N, Seita J, Hanna J, Murakami P, Jaenisch R, Weissleder R, Orkin SH, Weissman IL, Feinberg AP, Daley GQ (2010) Epigenetic memory in induced pluripotent stem cells. *Nature* **467:** 285-290

Kim K, Kim H, Lee D (2009) Site-specific modification of genome with cell-permeable Cre fusion protein in preimplantation mouse embryo. *Biochemical and biophysical research communications* **388:** 122-126

Kim KY, Hysolli E, Park IH (2011) Neuronal maturation defect in induced pluripotent stem cells from patients with Rett syndrome. *Proceedings of the National Academy of Sciences of the United States of America* **108:** 14169-14174

Kim S, Kim D, Cho SW, Kim J, Kim JS (2014) Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome research* **24:** 1012-1019

Kime C, Mandegar MA, Srivastava D, Yamanaka S, Conklin BR, Rand TA (2016) Efficient CRISPR/Cas9-Based Genome Engineering in Human Pluripotent Stem Cells. *Current protocols in human genetics / editorial board, Jonathan L Haines [et al]* **88:** Unit 21 24

Kinde B, Wu DY, Greenberg ME, Gabel HW (2016) DNA methylation in the gene body influences MeCP2-mediated gene repression. *PNAS* **113:** 15114-15119

Kishi N, Macklis JD (2004) MECP2 is progressively expressed in post-migratory neurons and is involved in neuronal maturation rather than cell fate decisions. *Molecular and cellular neurosciences* **27:** 306-321

Kitt CA, Wilcox BJ (1995) Preliminary evidence for neurodegenerative changes in the subtantia nigra of Rett syndrome. *Neuropediatrics* **26:** 114-118

Klose RJ, Bird AP (2004) MeCP2 behaves as an elongated monomer that does not stably associate with the Sin3a chromatin remodeling complex. *The Journal of biological chemistry* **279:** 46490-46496

Klose RJ, Sarraf SA, Schmiedeberg L, McDermott SM, Stancheva I, Bird AP (2005) DNA binding selectivity of MeCP2 due to a requirement for A/T sequences adjacent to methyl-CpG. *Mol Cell* **19:** 667-678

Kokura K, Kaul SC, Wadhwa R, Nomura T, Khan MM, Shinagawa T, Yasukawa T, Colmenares C, Ishii S (2001) The Ski protein family is required for MeCP2-mediated transcriptional repression. *The Journal of biological chemistry* **276:** 34115-34121

Komor AC, Badran AH, Liu DR (2017) CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. *Cell* **168:** 20-36

Kotin RM, Linden RM, Berns Kl (1992) Characterization of a preferred site on human chromosome 19q for integration of adeno-associated virus DNA by non-homologous recombination. *The EMBO journal* **11:** 5071-5078

Kotin RM, Siniscalco M, Samulski RJ, Zhu X, Hunter L, Laughlin CA, McLaughlin S, Muzycza N, Rocchi M, Berns KI (1990) Site-specific integration by adeno-associated virus. *PNAS*

Kriaucionis S, Bird A (2004) The major form of MeCP2 has a novel N-terminus generated by alternative splicing. *Nucleic acids research* **32:** 1818-1823

Kriaucionis S, Paterson A, Curtis J, Guy J, Macleod N, Bird A (2006) Gene expression analysis exposes mitochondrial abnormalities in a mouse model of Rett syndrome. *Molecular and cellular biology* **26:** 5033-5042

Krishna A, Biryukov M, Trefois C, Antony PM, Hussong R, Lin J, Heinäniemi M, Glusman G, Köglsberger S, Boyd O, Berg BHvd, Linke D, Huang D, Wang K, Hood L, Tholey A, Schneider R, Galas DJ, Balling R, May P (2014) Systems genomics evaluation of the SH-SY5Y neuroblastoma cell line as a model for Parkinson's disease. *BMC genomics*

Kruusvee V, Lyst MJ, Taylor C, Tarnauskaite Z, Bird AP, Cook AG (2017) Structure of the MeCP2-TBLR1 complex reveals a molecular basis for Rett syndrome and related disorders. *Proceedings of the National Academy of Sciences of the United States of America*

Kucukkal TG, Yang Y, Uvarov O, Cao W, Alexov E (2015) Impact of Rett Syndrome Mutations on MeCP2 MBD Stability. *Biochemistry* **54:** 6357-6368

Kuijlaars J, Oyelami T, Diels A, Rohrbacher J, Versweyveld S, Meneghello G, Tuefferd M, Verstraelen P, Detrez JR, Verschuuren M, De Vos WH, Meert T, Peeters PJ, Cik M, Nuydens R, Brone B, Verheyen A (2016) Sustained synchronized neuronal network activity in a human astrocyte co-culture system. *Scientific reports* **6:** 36529

Laccone FA (2012) MeCP2 protein therapy patent.

Laget S, Joulie M, Masson FL, Sasai N, Christians E, Pradhan S, Roberts RJ, Defossez P-A (2010) The Human Proteins MBD5 and MBD6 Associate with Heterochromatin but They Do Not Bind Methylated DNA. *PloS one* **5:** 1-11

Lagger S, Connelly JC, Schweikert G, Webb S, Selfridge J, Ramsahoye BH, Yu M, He C, Sanguinetti G, Sowers LC, Walkinshaw MD, Bird A (2017) MeCP2 recognizes cytosine methylated tri-nucleotide and di-nucleotide sequences to tune transcription in the mammalian brain. *PLoS Genet* **13:** e1006793

Lamonica JM, Kwon DY, Goffin D, Fenik P, Johnson BS, Cui Y, Guo H, Veasey S, Zhou Z (2017) Elevating expression of MeCP2 T158M rescues DNA binding and Rett syndrome-like phenotypes. *The Journal of clinical investigation* **127:** 1889-1904

LaSalle JM, Goldstine J, Balmer D, Greco CM (2001) Quantitative localization of heterogeneous methyl-CpG-binding protein 2 (MeCP2) expression phenotypes in normal and Rett syndrome brain by laser scanning cytometry. *Human molecular genetics* **10:** 1729–1740

Laurent L, Wong E, Li G, Huynh T, Tsirigos A, Ong CT, Low HM, Kin Sung KW, Rigoutsos I, Loring J, Wei CL (2010) Dynamic changes in the human methylome during differentiation. *Genome research* **20:** 320-331

Lawson-Yuen A, Liu D, Han L, Jiang ZI, Tsai GE, Basu AC, Picker J, Feng J, Coyle JT (2007) Ube3a mRNA and protein expression are not decreased in Mecp2R168X mutant mice. *Brain research* **1180:** 1-6

Lee G, Ramirez CN, Kim H, Zeltner N, Liu B, Radu C, Bhinder B, Kim YJ, Choi IY, Mukherjee-Clavin B, Djaballah H, Studer L (2012) Large-scale screening using familial dysautonomia induced pluripotent stem cells identifies compounds that rescue IKBKAP expression. *Nature biotechnology* **30:** 1244-1248

Leonard H, Cobb S, Downs J (2017) Clinical and biological progress over 50 years in Rett syndrome. *Nature reviews Neurology* **13:** 37-51

Lewis JD, Meehan RR, Henzel WJ, Maurer-Fogy I, Jeppesen P, Klein F, Bird A (1992) Purification, Sequence, and Cellular Localization of a Novel Chromosomal Protein That Binds to Methylated DNA. *Cell* **69**

Li D, Qiu Z, Shao Y, Chen Y, Guan Y, Liu M, Li Y, Gao N, Wang L, Lu X, Zhao Y, Liu M (2013a) Heritable gene targeting in the mouse and rat using a CRISPR-Cas system. *Nature biotechnology* **31**

Li E, Bestor TH, Jaenisch R (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69:** 915-926

Li L, Gao F, Wu S (2016) An episomal CRISPR/Cas9 system to derive vector-free gene modified mammalian cells. *Protein & cell*

Li W, Teng F, Li T, Zhou Q (2013b) Simultaneous generation and germline transmission of multiple gene mutations in rat using CRISPR-Cas systems. *Nature biotechnology* **31**

Li Y, Wang H, Muffat J, Cheng AW, Orlando DA, Loven J, Kwok SM, Feldman DA, Bateup HS, Gao Q, Hockemeyer D, Mitalipova M, Lewis CA, Vander Heiden MG, Sur M, Young RA, Jaenisch R (2013c) Global transcriptional and translational repression in human-embryonic-stem-cell-derived Rett syndrome neurons. *Cell stem cell* **13:** 446-458

Liang P, Xu Y, Zhang X, Ding C, Huang R, Zhang Z, Lv J, Xie X, Chen Y, Li Y, Sun Y, Bai Y, Songyang Z, Ma W, Zhou C, Huang J (2015a) CRISPR/Cas9-mediated gene editing in human tripronuclear zygotes. *Protein & cell*

Liang X, Potter J, Kumar S, Zou Y, Quintanilla R, Sridharan M, Carte J, Chen W, Roark N, Ranganathan S, Ravinder N, Chesnut JD (2015b) Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. *Journal of biotechnology* **208:** 44-53

Liao Y, Smyth GK, Shi W (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30:** 923-930

Lieber MR (2010) The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annual review of biochemistry* **79:** 181-211

Lin Q, Jo D, Gebre-Amlak KD, Ruley HE (2004) Enhanced cell-permeant Cre protein for site-specific recombination in cultured cells. *BMC biotechnology* **4:** 25

Lin S, Staahl BT, Alla RK, Doudna JA (2014) Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife* **3:** e04766

Linhoff MW, Garg SK, Mandel G (2015) A High-Resolution Imaging Approach to Investigate Chromatin Architecture in Complex Tissues. *Cell* **163:** 246-255

Lioy DT, Garg SK, Monaghan CE, Raber J, Foust KD, Kaspar BK, Hirrlinger PG, Kirchhoff F, Bissonnette JM, Ballas N, Mandel G (2011) A role for glia in the progression of Rett's syndrome. *Nature* **475:** 497-500

Lister R, Mukamel EA, Nery JR, Urich M, Puddifoot CA, Johnson ND, Lucero J, Huang Y, Dwork AJ, Schultz MD, Yu M, Tonti-Filippini J, Heyn H, Hu S, Wu JC, Rao A, Esteller M, He C, Haghighi FG, Sejnowski TJ, Behrens MM, Ecker JR (2013) Global epigenomic reconfiguration during mammalian brain development. *Science* **341:** 1237905

Liu H, Chen Y, Niu Y, Zhang K, Kang Y, Ge W, Liu X, Zhao E, Wang C, Lin S, Jing B, Si C, Lin Q, Chen X, Lin H, Pu X, Wang Y, Qin B, Wang F, Wang H, Si W, Zhou J, Tan T, Li T, Ji S, Xue Z, Luo Y, Cheng L, Zhou Q, Li S, Sun YE, Ji W (2014a) TALEN-Mediated Gene Mutagenesis in Rhesus and Cynomolgus Monkeys. *Cell stem cell* **14:** 323-328

Liu J, Gaj T, Yang Y, Wang N, Shui S, Kim S, Kanchiswamy CN, Kim J-S, Barbas CF (2015) Efficient delivery of nuclease proteins for genome editing in human stem cells and primary cells. *Nature protocols* **10:** 1842-1859

Liu Z, Brown A, Fisher D, Wu Y, Warren J, Cui X (2016) Tissue Specific Expression of Cre in Rat Tyrosine Hydroxylase and Dopamine Active Transporter-Positive Neurons. *PloS one* **11:** e0149379

Liu Z, Zhou X, Zhu Y, Chen ZF, Yu B, Wang Y, Zhang CC, Nie YH, Sang X, Cai YJ, Zhang YF, Zhang C, Zhou WH, Sun Q, Qiu Z (2014b) Generation of a monkey with MECP2 mutations by TALEN-based gene targeting. *Neuroscience bulletin* **30:** 381-386

Lombardi LM, Baker SA, Zoghbi HY (2015) MECP2 disorders: from the clinic to mice and back. *The Journal of clinical investigation* **125:** 2914-2923

Lorincz MC, Schubeler D, Groudine M (2001) Methylation-mediated proviral silencing is associated with MeCP2 recruitment and localized histone H3 deacetylation. *Molecular and cellular biology* **21:** 7913-7922

Lotharius J, Barg S, Wiekop P, Lundberg C, Raymon HK, Brundin P (2002) Effect of Mutant α-Synuclein on Dopamine Homeostasis in a New Human Mesencephalic Cell Line. *Journal of Biological Chemistry* **277:** 38884-38894

Lotharius J, Falsig J, van Beek J, Payne S, Dringen R, Brundin P, Leist M (2005) Progressive degeneration of human mesencephalic neuron-derived cells triggered by dopamine-dependent oxidative stress is dependent on the mixed-lineage kinase pathway. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **25:** 6329-6342

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* **15:** 550

Luikenhuis S, Giacometti E, Beard CF, Jaenisch R (2004) Expression of MeCP2 in postmitotic neurons rescues Rett syndrome in mice. *Proceedings of the National Academy of Sciences of the United States of America* **101:** 6033-6038

Lunyak VV, Burgess R, Prefontaine GG, Nelson C, Sze S-H, Chenoweth J, Schwartz P, Pevzner PA, Glass C, Mandel G, Rosenfeld MG (2002) Corepressor-Dependent Silencing of Chromosomal Regions Encoding Neuronal Genes. *Science* **298:** 1747-1752

Luthman H, Magnusson G (1983) HIgh efficiency polyoms DNA transfecion of chloroquine treated cels. *NAR* **11**

Lyst MJ, Bird A (2015) Rett syndrome: a complex disorder with simple roots. *Nature reviews Genetics* **16:** 261–275

Lyst MJ, Connelly J, Merusi C, Bird A (2016) Sequence specific DNA binding by AT-hook motifs in MeCP2. *FEBS letters*

Lyst MJ, Ekiert R, Ebert DH, Merusi C, Nowak J, Selfridge J, Guy J, Kastan NR, Robinson ND, de Lima Alves F, Rappsilber J, Greenberg ME, Bird A (2013) Rett syndrome mutations abolish the interaction of MeCP2 with the NCoR/SMRT co-repressor. *Nature neuroscience* **16:** 898-902

M.L. Bauman M, T.L. Kemper M, Arin DM (1995) Pervasive neuroanatomic abnormalities of the brain in three cases of Rett's syndrome. *Neurology*

Makarova KS, Grishin NV, Shabalina SA, Wolf YI, Koonin EV (2006) A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biology direct* **1:** 7

Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM (2013) RNA-guided human genome engineering via Cas9. *Science* **339:** 823-826

Manno CS, Pierce GF, Arruda VR, Glader B, Ragni M, Rasko JJ, Ozelo MC, Hoots K, Blatt P, Konkle B, Dake M, Kaye R, Razavi M, Zajko A, Zehnder J, Rustagi PK, Nakai H, Chew A, Leonard D, Wright JF, Lessard RR, Sommer JM, Tigges M, Sabatino D, Luk A, Jiang H, Mingozzi F, Couto L, Ertl HC, High KA, Kay MA (2006) Successful transduction of liver in hemophilia by AAV-Factor IX and limitations imposed by the host immune response. *Nature medicine* **12:** 342-347

Marchetto MC, Carromeu C, Acab A, Yu D, Yeo GW, Mu Y, Chen G, Gage FH, Muotri AR (2010) A model for neural development and treatment of Rett syndrome using human induced pluripotent stem cells. *Cell* **143:** 527-539

Marshak S, Meynard MM, De Vries YA, Kidane AH, Cohen-Cory S (2012) Cell-autonomous alterations in dendritic arbor morphology and connectivity induced by overexpression of MeCP2 in Xenopus central neurons in vivo. *PloS one* **7:** e33153

Martinowich K, Hattori D, Wu H, Fouse S, He F, Hu Y, Fan G, Sun YE (2003) DNA methylation-related chromatin remodeling in activity-dependent BDNF gene regulation. *Science* **302:** 890-893

Maruyama T, Dougan SK, Truttmann MC, Bilate AM, Ingram JR, Ploegh HL (2015) Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nature biotechnology* **33:** 538-542

Maunakea AK, Chepelev I, Cui K, Zhao K (2013) Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. *Cell research* **23:** 1256-1269

Maunakea AK, Nagarajan RP, Bilenky M, Ballinger TJ, D'Souza C, Fouse SD, Johnson BE, Hong C, Nielsen C, Zhao Y, Turecki G, Delaney A, Varhol R, Thiessen N, Shchors K, Heine VM, Rowitch DH, Xing X, Fiore C, Schillebeeckx M, Jones SJ, Haussler D, Marra MA, Hirst M, Wang T, Costello JF (2010) Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* **466:** 253-257

McCarty DM, Young SM, Jr., Samulski RJ (2004) Integration of adeno-associated virus (AAV) and recombinant AAV vectors. *Annual review of genetics* **38:** 819-845

McClure C, Cole KL, Wulff P, Klugmann M, Murray AJ (2011) Production and titering of recombinant adeno-associated viral vectors. *Journal of visualized experiments : JoVE***:** e3348

McGhee J, Ginder G (1979) Specific DNA methylation sites in the vicinity of the chicken beta-globin genes. *Nature***:** 419-420

McGowan H, Pang ZP (2015) Regulatory functions and pathological relevance of the MECP2 3'UTR in the central nervous system. *Cell regeneration* **4:** 9

McGraw CM, Samaco RC, Zoghbi HY (2011) Adult neural function requires MeCP2. *Science* **333:** 186

Meehan RR, Lewis JD, Bird AP (1992) Characterization of MeCP2, a vertebrate DNA binding protein with affinity for methylated DNA. *Nucleic acids research* **20**

Meehan RR, Lewis JD, McKay S, Kleiner EL, Bird AP (1989) Identification of a Mammalian Protein That Binds Specifically to DNA Containing Methylated CpGs. *Cell* **58:** 499-507

Mellen M, Ayata P, Dewell S, Kriaucionis S, Heintz N (2012) MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell* **151:** 1417-1430

Meng X, Wang W, Lu H, He L-j, Chen W, Chao E, Fiorotto ML, Tang B, Herrera JA, Seymour ML, Neul JL, Pereira FA, Tang J, Xue M, Zoghbi HY (2016) Manipulations of MeCP2 in glutamatergic neurons highlight their contributions to Rett and other neurological disorders. *eLIFE*

Merkert S, Martin U (2016) Site-Specific Genome Engineering in Human Pluripotent Stem Cells. *International journal of molecular sciences* **17**

Mi Z, Mai J, Lu X, Robbins PD (2000) Characterization of a class of cationic peptides able to facilitate efficient protein transduction in vitro and in vivo. *Molecular therapy : the journal of the American Society of Gene Therapy* **2:** 339-347

Mironov SL, Skorova E, Hartelt N, Mironova LA, Hasan MT, Kugler S (2009) Remodelling of the respiratory network in a mouse model of Rett syndrome depends on brain-derived neurotrophic factor regulated slow calcium buffering. *The Journal of physiology* **587:** 2473-2485

Mitchell DJ, Kim DT, Steinman L, Fathman CG, Rothbard JB (2000) Polyarginine enters cells more efficiently than other polycationic homopolymers. *J Peptide Res* **56:** 318-325

Miyaoka Y, Berman JR, Cooper SB, Mayerl SJ, Chan AH, Zhang B, Karlin-Neumann GA, Conklin BR (2016) Systematic quantification of HDR and NHEJ reveals effects of locus, nuclease, and cell type on genome-editing. *Scientific reports* **6:** 23549

Mnatzakanian GN, Lohi H, Munteanu I, Alfred SE, Yamada T, MacLeod PJ, Jones JR, Scherer SW, Schanen NC, Friez MJ, Vincent JB, Minassian BA (2004) A previously unidentified MECP2 open reading frame defines a new protein isoform relevant to Rett syndrome. *Nature genetics* **36:** 339-341

Mohandas T, Sparkes R, Shapiro L (1981) Reactivation of an Inactive Human X Chromosome: evidence for X Inactivation by DNA Methylation. *Science* **211:** 393-396

Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Soria E (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of molecular evolution* **60:** 174-182

Mojica FJM, Diez-Villasenor C, Soria E, Juez G (2000) Biological significance of a family of regularly space repeats in the genomes of Archaea, Bacteria and mitochondria. *Molecular Microbiology* **36:** 244-246

Moore LD, Le T, Fan G (2013) DNA methylation and its basic function. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology* **38:** 23-38

Moretti P, Levenson JM, Battaglia F, Atkinson R, Teague R, Antalffy B, Armstrong D, Arancio O, Sweatt JD, Zoghbi HY (2006) Learning and memory and synaptic plasticity are impaired in a mouse model of Rett syndrome. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **26:** 319-327

Mungenast AE, Siegert S, Tsai LH (2016) Modeling Alzheimer's disease with human induced pluripotent stem (iPS) cells. *Molecular and cellular neurosciences* **73:** 13-31

Muotri AR, Marchetto MC, Coufal NG, Oefner R, Yeo G, Nakashima K, Gage FH (2010) L1 retrotransposition in neurons is modulated by MeCP2. *Nature* **468:** 443-446

Murphy SL, Li H, Zhou S, Schlachterman A, High K (2008) Prolonged Susceptibility to Antibody-mediated Neutralization for Adeno-associated Vectors Targeted to the Liver. *Molecular Therapy* **16:** 138-145

Nageshappa S, Carromeu C, Trujillo CA, Mesci P, Espuny-Camacho I, Pasciuto E, Vanderhaeghen P, Verfaillie CM, Raitano S, Kumar A, Carvalho CM, Bagni C, Ramocki MB, Araujo BH, Torres LB, Lupski JR, Van Esch H, Muotri AR (2015) Altered neuronal network and rescue in a human MECP2 duplication model. *Molecular psychiatry*

Nakade S, Tsubota T, Sakane Y, Kume S, Sakamoto N, Obara M, Daimon T, Sezutsu H, Yamamoto T, Sakuma T, Suzuki KT (2014) Microhomology-mediated end-joining-dependent integration of donor DNA in cells and animals using TALENs and CRISPR/Cas9. *Nature communications* **5:** 5560

Nakai H, Yant SR, Storm TA, Fuess S, Meuse L, Kay MA (2001) Extrachromosomal recombinant adeno-associated virus vector genomes are primarily responsible for stable liver transduction in vivo. *Journal of virology* **75:** 6969-6976

Nakajima J, Okamoto N, Tohyama J, Kato M, Arai H, Funahashi O, Tsurusaki Y, Nakashima M, Kawashima H, Saitsu H, Matsumoto N, Miyake N (2015) De novo EEF1A2 mutations in patients with characteristic facial features, intellectual disability, autistic behaviors and epilepsy. *Clinical genetics* **87:** 356-361

Nan X, Campoy FJ, Bird A (1997) MeCP2 Is a Transcriptional Repressor with Abundant Binding Sites in Genomic Chromatin. *Cell*

Nan X, Hou J, Maclean A, Nasir J, Lafuente MJ, Shu X, Kriaucionis S, Bird A (2007) Interaction between chromatin proteins MECP2 and ATRX is disrupted by mutations that cause inherited mental retardation. *Proceedings of the National Academy of Sciences of the United States of America* **104:** 2709-2714

Nan X, Meehan RR, Bird AP (1993) Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. *Nucleic acids research* **21**

Nan X, Ng H-H, Johnson CA, Laherty CD, Turner BM, Eisenman RN, Bird A (1998) Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deactylase comlex. *Nature* **393**

Nan X, Tate P, Li E, Bird A (1996) DNA methylation specifies chromosomal localization of MeCP2. *Molecular and cellular biology* **16**

Natsume T, Kiyomitsu T, Saga Y, Kanemaki Masato T (2016) Rapid Protein Depletion in Human Cells by Auxin-Inducible Degron Tagging with Short Homology Donors. *Cell reports* **15:** 210-218

Nectoux J, Fichou Y, Rosas-Vargas H, Cagnard N, Bahi-Buisson N, Nusbaum P, Letourneur F, Chelly J, Bienvenu T (2010) Cell cloning-based transcriptome analysis in Rett patients: relevance to the pathogenesis of Rett syndrome of new human MeCP2 target genes. *Journal of cellular and molecular medicine* **14:** 1962-1974

Neul JL, Kaufmann WE, Glaze DG, Christodoulou J, Clarke AJ, Bahi-Buisson N, Leonard H, Bailey ME, Schanen NC, Zappella M, Renieri A, Huppke P, Percy AK, RettSearch C (2010) Rett syndrome: revised diagnostic criteria and nomenclature. *Ann Neurol* **68:** 944-950

Ng H-H, Jeppeson P, Bird A (2000) Active Repression of Methylated Genes by the Chromosomal Protein MBD1. *Molecular and cellular biology* **20:** 1394-1406

Nguyen S, Meletis K, Fu D, Jhaveri S, Jaenisch R (2007) Ablation of de novo DNA methyltransferase Dnmt3a in the nervous system leads to neuromuscular defects and shortened lifespan. *Developmental dynamics : an official publication of the American Association of Anatomists* **236:** 1663-1676

Nikitina T, Ghosh RP, Horowitz-Scherer RA, Hansen JC, Grigoryev SA, Woodcock CL (2007a) MeCP2-chromatin interactions include the formation of chromatosome-like structures and are altered in mutations causing Rett syndrome. *The Journal of biological chemistry* **282:** 28237-28245

Nikitina T, Shi X, Ghosh RP, Horowitz-Scherer RA, Hansen JC, Woodcock CL (2007b) Multiple modes of interaction between the methylated DNA binding protein MeCP2 and chromatin. *Molecular and cellular biology* **27:** 864-877

Niu Y, Shen B, Cui Y, Chen Y, Wang J, Wang L, Kang Y, Zhao X, Si W, Li W, Xiang AP, Zhou J, Guo X, Bi Y, Si C, Hu B, Dong G, Wang H, Zhou Z, Li T, Tan T, Pu X, Wang F, Ji S, Zhou Q, Huang X, Ji W, Sha J (2014a) Generation of gene-modified cynomolgus monkey via Cas9/RNA-mediated gene targeting in one-cell embryos. *Cell* **156:** 836-843

Niu Y, Shen B, Cui Y, Chen Y, Wang J, Wang L, Kang Y, Zhao X, Si W, Li W, Xiang AP, Zhou J, Guo X, Bi Y, Si C, Hu B, Dong G, Wang H, Zhou Z, Li T, Tan T, Pu X, Wang F, Ji S, Zhou Q, Huang X, Ji W, Sha J (2014b) Generation of Gene-Modified Cynomolgus Monkey via Cas9/RNA-Mediated Gene Targeting in One-Cell Embryos. *Cell*

Nojima T, Gomes T, Grosso AR, Kimura H, Dye MJ, Dhir S, Carmo-Fonseca M, Proudfoot NJ (2015) Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. *Cell* **161:** 526-540

Nowak D, Hofmann WK, Koeffler HP (2009) Genome-wide Mapping of Copy Number Variations Using SNP Arrays. *Transfus Med Hemother* **36:** 246-251

Nuber UA, Kriaucionis S, Roloff TC, Guy J, Selfridge J, Steinhoff C, Schulz R, Lipkowitz B, Ropers HH, Holmes MC, Bird A (2005) Up-regulation of glucocorticoid-regulated genes in a mouse model of Rett syndrome. *Human molecular genetics* **14:** 2247-2256

Odawara A, Saitoh Y, Alhebshi AH, Gotoh M, Suzuki I (2014) Long-term electrophysiological activity and pharmacological response of a human induced pluripotent stem cell-derived neuron and astrocyte co-culture. *Biochemical and biophysical research communications* **443:** 1176-1181

Ohi Y, Qin H, Hong C, Blouin L, Polo JM, Guo T, Qi Z, Downey SL, Manos PD, Rossi DJ, Yu J, Hebrok M, Hochedlinger K, Costello JF, Song JS, Ramalho-Santos M (2011) Incomplete DNA methylation underlies a transcriptional memory of somatic cells in human iPS cells. *Nature cell biology* **13:** 541-549

Okada CY, Rechsteiner M (1982) Introduction of macromolecules by into cultured mammalian cells by osmotic lysis of pinocytic vesicles. *Cell* **29:** 33-41

Okano M, Bell DW, Haber DA, Li E (1999) DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. *Cell* **99:** 247-257

Orii KE, Lee Y, Kondo N, McKinnon PJ (2006) Selective utilization of nonhomologous end-joining and homologous recombination DNA repair pathways during nervous system development. *Proceedings of the National Academy of Sciences of the United States of America* **103:** 10017-10022

Padovan-Merhar O, Nair GP, Biaesch AG, Mayer A, Scarfone S, Foley SW, Wu AR, Churchman LS, Singh A, Raj A (2015) Single mammalian cells compensate for differences in cellular volume and DNA copy number through independent global transcriptional mechanisms. *Mol Cell* **58:** 339-352

Paquet D, Kwart D, Chen A, Sproul A, Jacob S, Teo S, Olsen KM, Gregg A, Noggle S, Tessier-Lavigne M (2016) Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. *Nature* **533:** 125-129

Patterson KC, Hawkins VE, Arps KM, Mulkey DK, Olsen ML (2016) MeCP2 deficiency results in robust Rett-like behavioural and motor deficits in male and female rats. *Human molecular genetics* **25:** 5514-5515

Paul G, Christophersen NS, Raymon H, Kiaer C, Smith R, Brundin P (2007) Tyrosine hydroxylase expression is unstable in a human immortalized mesencephalic cell line--studies in vitro and after intracerebral grafting in vivo. *Molecular and cellular neurosciences* **34:** 390-399

Pearson EC, Bates DL, Prospero TD, Thomas JO (1984) Neuronal nuclei and glial nuclei from mammalian cerebral cortex: Nucleosome repeat lengths, DNA contents and H1 contents. *Eur J Biochem* **144:** 353 - 360

Peitz M, Pfannkuche K, Rajewsky K, Edenhofer F (2002) Ability of the hydrophobic FGF and basic TAT peptides to promote cellular uptake of recombinant Cre recombinase: a tool for efficient genetic engineering of mammalian genomes. *Proceedings of the National Academy of Sciences of the United States of America* **99:** 4489-4494

Pelka GJ, Watson CM, Christodoulou J, Tam PP (2005) Distinct expression profiles of Mecp2 transcripts with different lengths of 3'UTR in the brain and visceral organs during mouse development. *Genomics* **85:** 441-452

Penaud-Budloo M, Le Guiner C, Nowrouzi A, Toromanoff A, Cherel Y, Chenuaud P, Schmidt M, von Kalle C, Rolling F, Moullier P, Snyder RO (2008) Adeno-associated virus vector genomes persist as episomal chromatin in primate muscle. *Journal of virology* **82:** 7875-7885

Philpott NJ, Giraud-Wali C, Dupuis C, Gomos J, Hamilton H, Berns KI, Falck-Pedersen E (2002) Efficient Integration of Recombinant Adeno-Associated Virus DNA Vectors Requires a p5-rep Sequence in cis. *Journal of virology* **76:** 5411-5421

Pigoni M, Wanngren J, Kuhn PH, Munro KM, Gunnersen JM, Takeshima H, Feederle R, Voytyuk I, De Strooper B, Levasseur MD, Hrupka BJ, Muller SA, Lichtenthaler SF (2016) Seizure protein 6 and its homolog seizure 6-like protein are physiological substrates of BACE1 in neurons. *Molecular neurodegeneration* **11:** 67

Pitcher MR, Herrera JA, Buffington SA, Kochukov MY, Merritt JK, Fisher AR, Schanen NC, Costa-Mattioli M, Neul JL (2015) Rett syndrome like phenotypes in the R255X Mecp2 mutant mouse are rescued by MECP2 transgene. *Human molecular genetics* **24:** 2662-2672

Platt RJ, Chen S, Zhou Y, Yim MJ, Swiech L, Kempton HR, Dahlman JE, Parnas O, Eisenhaure TM, Jovanovic M, Graham DB, Jhunjhunwala S, Heidenreich M, Xavier RJ, Langer R, Anderson DG, Hacohen N, Regev A, Feng G, Sharp PA, Zhang F (2014) CRISPR-Cas9 Knockin Mice for Genome Editing and Cancer Modeling. *Cell* **159:** 440-455

Plaza Reyes A, Lanner F (2017) Towards a CRISPR view of early human development: applications, limitations and ethical concerns of genome editing in human embryos. *Development* **144:** 3-7

Pomp O, Dreesen O, Leong DF, Meller-Pomp O, Tan TT, Zhou F, Colman A (2011) Unexpected X chromosome skewing during culture and reprogramming of human somatic cells can be alleviated by exogenous telomerase. *Cell stem cell* **9:** 156-165

Porteus MH, Baltimore D (2003) Chimeric nucleases stimulate gene targeting in human cells. *Science* **300:** 763

Porteus MH, Cathomen T, Weitzman MD, Baltimore D (2003) Efficient gene targeting mediated by adeno-associated virus and DNA double-strand breaks. *Molecular and cellular biology* **23:** 3558-3565

Pourcel C, Salvignol G, Vergnaud G (2005) CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* **151:** 653-663

Prokhortchouk A, Hendrich B, Jørgensen H, Ruzov A, Wilm M, Georgiev G, Bird A, Prokhortchouk E (2002) The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. *Genes & development* **15:** 1613–1618

Quaderi NA, Meehan RR, Tate PH, Cross SH, Bird AP, Chatterjee A, Herman GE, Brown SDM (1994) Genetic and Physical Mapping of a Gene Encoding a Methyl CpG Binding Protein, Mecp2, to the Mouse X Chromosome. *Genomics* **22:** 648-651

R.J.Samulski, X.Zhu, X.Xiao, J.D.Brook, D.E.Housman, N.Epstein, L.A.Hunter (1991) Targeted integration of adeno-associated virus (AAV) into human chromosome 19. *The EMBO journal* **10:** 3941-3950

Ramakrishna S, Kwaku Dad AB, Beloor J, Gopalappa R, Lee SK, Kim H (2014) Gene disruption by cell-penetrating peptide-mediated delivery of Cas9 protein and guide RNA. *Genome research* **24:** 1020-1027

Ramocki MB, Tavyev YJ, Peters SU (2010) The MECP2 duplication syndrome. *American journal of medical genetics Part A* **152A:** 1079-1088

Ramsahoye BH, Biniszkiewicz D, Lyko F, Clark V, Bird AP, Jaenisch R (2000) Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *PNAS* **97:** 5237-5342

Ran FA, Hsu PD, Lin CY, Gootenberg JS, Konermann S, Trevino AE, Scott DA, Inoue A, Matoba S, Zhang Y, Zhang F (2013a) Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* **154:** 1380-1389

Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F (2013b) Genome engineering using the CRISPR-Cas9 system. *Nature protocols* **8:** 2281-2308

Rastegar M, Hotta A, Pasceri P, Makarem M, Cheung AYL, Elliot S, Park KJ, Adachi M, Jones FS, Clarke ID, Dirks P, Ellis J (2009) MECP2 Isoform-Specific Vectors with Regulated Expression for Rett Syndrome Gene Therapy. *PloS one* **4**

Rauch TA, Wu X, Zhong X, Riggs AD, Pfeifer GP (2009) A human B cell methylome at 100-base pair resolution. *Proceedings of the National Academy of Sciences of the United States of America* **106:** 671-678

Ravn K, Nielsen JB, Uldall P, Hansen FJ, Schwartz M (2003) No correlation between phenotype and genotype in boys with truncating *MECP2* mutation. *Journal of medical genetics* **40:** 1-5

Reichwald K, Thiesen J, Wiehe T, Weitzel J, Stratling WH, Kioschis P, Poustka A, Rosenthal A, Platzer M (2000) Comparative sequence analysis of the MECP2-locus in human and mouse reveals new transcribed regions. *Mammalian Genome* **11:** 182-190

Rekaik H, Blaudin de The FX, Prochiantz A, Fuchs J, Joshi RL (2015) Dissecting the role of Engrailed in adult dopaminergic neurons--Insights into Parkinson disease pathogenesis. *FEBS letters* **589:** 3786-3794

Rett A (2016) On a remarkable syndrome of cerebral atrophy associated with hyperammonaemia in childhood. *Wiener medizinische Wochenschrift* **166:** 322-324

Richardson CD, Kazane KR, Feng SJ, Bray NL, Schäfer AJ, Floor S, Corn JE. (2017) CRISPR-Cas9 genome editing in human cells works via the Fanconi Anemia pathway. BioRxiv.

Richardson CD, Ray GJ, DeWitt MA, Curie GL, Corn JE (2016) Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nature biotechnology* **34:** 339-344

Rietveld LEG, Caldenhoven E, Stunnenberg HG (2002) In vivo repression of an erythriod-specific gene by distinct corepressor complexes. *The EMBO journal* **21:** 1389-1397

Riggs AD (1975) X inactivation, differentiation, and DNA methylation. *Cytogenet Cell Genet* **14:** 9-25

Rodrigues DC, Kim DS, Yang G, Zaslavsky K, Ha KC, Mok RS, Ross PJ, Zhao M, Piekna A, Wei W, Blencowe BJ, Morris Q, Ellis J (2016) MECP2 Is Post-transcriptionally Regulated during Human Neurodevelopment by Combinatorial Action of RNA-Binding Proteins and miRNAs. *Cell reports* **17:** 720-734

Roloff TC, Ropers HH, Nuber UA (2003) Comparative study of methyl-CpG-binding domain proteins. *BMC genomics* **4:** 1-9

Ross PD, Guy J, Selfridge J, Kamal B, Bahey N, Tanner KE, Gillingwater TH, Jones RA, Loughrey CM, McCarroll CS, Bailey ME, Bird A, Cobb S (2016) Exclusive expression of MeCP2 in the nervous system distinguishes between brain and peripheral Rett syndrome-like phenotypes. *Human molecular genetics* **25:** 4389–4404

Rountree MR, Selker EU (1997) DNA methylation inhibits elongation but not initiation of transcription in Neurospora crassa. *Genes & development* **11:** 2383-2395

Rube HT, Lee W, Hejna M, Chen H, Yasui DH, Hess JF, LaSalle JM, Song JS, Gong Q (2016) Sequence features accurately predict genome-wide MeCP2 binding in vivo. *Nature communications* **7:** 11025

Russel DW, Hirata RK (1998) Human Gene Targeting by Viral Vectors. *Nature genetics* **18**

Samaco RC, Mandel-Brehm C, Chao HT, Ward CS, Fyffe-Maricich SL, Ren J, Hyland K, Thaller C, Maricich SM, Humphreys P, Greer JJ, Percy A, Glaze DG, Zoghbi HY, Neul JL (2009) Loss of MeCP2 in aminergic neurons causes cell-autonomous defects in neurotransmitter synthesis

and specific behavioral abnormalities. *Proceedings of the National Academy of Sciences of the United States of America* **106:** 21966-21971

Samulski RJ, Muzyczka N (2014) AAV-Mediated Gene Therapy for Research and Therapeutic Purposes. *Annual review of virology* **1:** 427-451

Sandor C, Robertson P, Lang C, Heger A, Booth H, Vowles J, Witty L, Bowden R, Hu M, Cowley SA, Wade-Martins R, Webber C (2017) Transcriptomic profiling of purified patient-derived dopamine neurons identifies convergent perturbations and therapeutics for Parkinson's disease. *Human molecular genetics* **26:** 552-566

Sanjana NE, Shalem O, Zhang F (2014) Improved vectors and genome-wide libraries for CRISPR screening. *Nature methods* **11:** 783-784

Schanen NC, Kurczynski WT, Brunelle D, Woodcock MM, Dure SD, Percy KA (1998) Neonatal Encephalopathy in Two Boys in Families With Recurrent Rett Syndrome. *Journal of Child Neurology* **13:** 229-231

Schanen NC, Roth Dahle EJ, Capozzoli F, Holm VA, Zoghbi HY, Francke U (1997) A New Rett Syndrome Family Consistent with X-Linked Inheritance Expands the X Chromosome Exclusion Map. *The American Journal of Human Genetics* **61:** 634-641

Schildknecht S, Karreman C, Pöltl D, Efrémova L, Kullmann C, Gutbier S, Krug A, Scholz D, Gerding HR, Leist M (2013) Generation of Genetically-Modified Human Differentiated Cells for Toxicological Tests and the Study of Neurodegenerative Diseases. *Altex*

Schlackow M, Nojima T, Gomes T, Dhir A, Carmo-Fonseca M, Proudfoot NJ (2017) Distinctive Patterns of Transcription and RNA Processing for Human lincRNAs. *Mol Cell* **65:** 25-38

Schmiedeberg L, Skene P, Deaton A, Bird A (2009) A Temporal Threshold for Formaldehyde Crosslinking and Fixation. *PloS one* **4**

Schnepp BC, Clark KR, Klemanski DL, Pacak CA, Johnson PR (2003) Genetic Fate of Recombinant Adeno-Associated Virus Vector Genomes in Muscle. *Journal of virology* **77:** 3495-3504

Schnepp BC, Jensen RL, Chen CL, Johnson PR, Clark KR (2005) Characterization of adeno-associated virus genomes isolated from human tissues. *Journal of virology* **79:** 14793-14803

Scholz D, Poltl D, Genewsky A, Weng M, Waldmann T, Schildknecht S, Leist M (2011) Rapid, complete and large-scale generation of post-mitotic neurons from the human LUHMES cell line. *Journal of neurochemistry* **119:** 957-971

Schubeler D (2015) Function and information content of DNA methylation. *Nature* **517:** 321-326

Schule B, Li HH, Fisch-Kohl C, Purmann C, Francke U (2007) DLX5 and DLX6 expression is biallelic and not modulated by MeCP2 deficiency. *American journal of human genetics* **81:** 492-506

Schumann K, Lin S, Boyer E, Simeonov DR, Subramaniam M, Gate RE, Haliburton GE, Ye CJ, Bluestone JA, Doudna JA, Marson A (2015) Generation of knock-in primary human T cells using Cas9 ribonucleoproteins. *Proceedings of the National Academy of Sciences of the United States of America* **112:** 10437-10442

Schwalb B, Michel M, Zacher B, Frühauf K, Demel C, Tresch A, Gagneur J, Cramer P (2016) TT-seq maps the human transient transcriptome. *Science* **362**

Schwarze SR, Ho A, Vocero-Akbani A, Dowdy SF (1999) In Vivo Protein Transduction: Delivery of a Biologically Active Protein into the Mouse. *Science* **285**

Shah RR, Bird AP (2017) MeCP2 mutations: progress towards understanding and treating Rett syndrome. *Genome medicine* **9:** 17

Shahbazian MD, Antalffy B, Armstrong DL, Zoghbi HY (2002a) Insight into Rett syndrome: MeCP2 levels display tissue- and cell-specific differences and correlate with neuronal maturation. *Human molecular genetics*

Shahbazian MD, Young JI, Yuva-Paylor LA, Spencer CM, Antalffy BA, Noebels JL, Armstrong DL, Paylor R, Zoghbi HY (2002b) Mice with Truncated MeCP2 Recapitulate Many Rett Syndrome Features and Display Hyperacetylation of Histone H3. *Neuron*

Sheridan C (2017) CRISPR therapeutics push into human testing. *Nat Biotech* **3:** 3-5

Shin S, Vemuri M (2010) Culture and Differentiation of Human Neural Stem Cells. *Protocols for Neural Cell Culture*

Shrivastav M, De Haro LP, Nickoloff JA (2008) Regulation of DNA double-strand break repair pathway choice. *Cell research* **18:** 134-147

Singleton MK, Gonzales ML, Leung KN, Yasui DH, Schroeder DI, Dunaway K, LaSalle JM (2011) MeCP2 is required for global heterochromatic and nucleolar changes during activity-dependent neuronal maturation. *Neurobiology of disease* **43:** 190-200

Sirianni N, Naidu S, Pereira J, Pillotto RF, Hoffman EP (1998) Rett syndrome: confirmation of X-linked dominant inheritance, and localization of the gene to Xq28. *American journal of human genetics* **63:** 1552-1558

Skene PJ, Illingworth RS, Webb S, Kerr AR, James KD, Turner DJ, Andrews R, Bird AP (2010) Neuronal MeCP2 is expressed at near histone-octamer levels and globally alters the chromatin state. *Mol Cell* **37:** 457-468

Smith ZD, Chan MM, Mikkelsen TS, Gu H, Gnirke A, Regev A, Meissner A (2012) A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature* **484:** 339-344

Smithies O, Gregg RG, Boggs SS, Koralewski MA, Kucherlapati RS (1985) Insertion of DNA sequences into the human chromosomal β-globin locus by homologous recombination. *Nature* **317:** 230-234

Sorek R, Kunin V, Hugenholtz P (2008) CRISPR — a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat Rev Microbiology* **6:** 181-186

Staahl BT, Benekareddy M, Coulon-Bainier C, Banfal AA, Floor SN, Sabo JK, Urnes C, Munares GA, Ghosh A, Doudna JA (2017) Efficient genome editing in the mouse brain by local delivery of engineered Cas9 ribonucleoprotein complexes. *Nature biotechnology*

Stancheva I, Collins AL, Van den Veyver IB, Zoghbi H, Meehan RR (2003) A Mutant Form of MeCP2 Protein Associated with Human Rett Syndrome Cannot Be Displaced from Methylated DNA by Notch in Xenopus Embryos. *Molecular Cell* **12:** 425-435

Stein JL, de la Torre-Ubieta L, Tian Y, Parikshak NN, Hernandez IA, Marchetto MC, Baker DK, Lu D, Hinman CR, Lowe JK, Wexler EM, Muotri AR, Gage FH, Kosik KS, Geschwind DH (2014) A quantitative framework to evaluate modeling of cortical development by neural stem cells. *Neuron* **83:** 69-86

Steinberg GK, Kondziolka D, Wechsler LR, Lunsford LD, Coburn ML, Billigen JB, Kim AS, Johnson JN, Bates D, King B, Case C, McGrogan M, Yankee EW, Schwartz NE (2016) Clinical Outcomes of Transplanted Modified Bone Marrow-Derived Mesenchymal Stem Cells in Stroke: A Phase 1/2a Study. *Stroke* **47:** 1817-1824

Su D, Cha YM, West AE (2012) Mutation of MeCP2 alters transcriptional regulation of select immediate-early genes. *Epigenetics : official journal of the DNA Methylation Society* **7:** 146-154

Sun Y, Pollard S, Conti L, Toselli M, Biella G, Parkin G, Willatt L, Falk A, Cattaneo E, Smith A (2008) Long-term tripotent differentiation capacity of human neural stem (NS) cells in adherent culture. *Molecular and cellular neurosciences* **38:** 245-258

Suzuki K, Tsunekawa Y, Hernandez-Benitez R, Wu J, Zhu J, Kim EJ, Hatanaka F, Yamamoto M, Araoka T, Li Z, Kurita M, Hishida T, Li M, Aizawa E, Guo S, Chen S, Goebl A, Soligalla RD, Qu J, Jiang T, Fu X, Jafari M, Esteban CR, Berggren WT, Lajara J, Nunez-Delicado E, Guillen P, Campistol JM, Matsuzaki F, Liu GH, Magistretti P, Zhang K, Callaway EM, Zhang K, Belmonte JC (2016) In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nature*

Swiech L, Heidenreich M, Banerjee A, Habib N, Li Y, Trombetta J, Sur M, Zhang F (2015) In vivo interrogation of gene function in the mammalian brain using CRISPR-Cas9. *Nature biotechnology* **33:** 102-106

Szulwach KE, Li X, Smrt RD, Li Y, Luo Y, Lin L, Santistevan NJ, Li W, Zhao X, Jin P (2010) Cross talk between microRNA and epigenetic regulation in adult neurogenesis. *The Journal of cell biology* **189:** 127-141

Tang X, Kim J, Zhou L, Wengert E, Zhang L, Wu Z, Carromeu C, Muotri AR, Marchetto MC, Gage FH, Chen G (2016) KCC2 rescues functional deficits in human neurons derived from patients with Rett syndrome. *Proceedings of the National Academy of Sciences of the United States of America*

Tang X, Zhou L, Wagner AM, Marchetto MCN, Muotri AR, Gage FH, Chen G (2013) Astroglial cells regulate the developmental timeline of human neurons differentiated from induced pluripotent stem cells. *Stem Cell Research* **11:** 743-757

Tarquinio DC, Hou W, Neul JL, Kaufmann WE, Glaze DG, Motil KJ, Skinner SA, Lee HS, Percy AK (2015) The Changing Face of Survival in Rett Syndrome and MECP2-Related Disorders. *Pediatric neurology* **53:** 402-411

Taylor SM, Jones PA (1979) Multiple new phenotypes induced in 10T1/2 and 3T3 cells treated with 5-aza-cytidine. *Cell* **17:** 771-779

Tebas P, Stein D, Tang WW, Frank I, Wang SQ, Lee G, Spratt SK, Surosky RT, Giedlin MA, Nichol G, Holmes MC, Gregory PD, Ando DG, Kalos M, Collman RG, Binder-Scholl G, Plesa G, Hwang W-T, Levine BL, June CH (2014) Gene Editing of CCR5 in Autologous CD4 T Cells of Persons Infected with HIV. *New England Journal of Medicine* **370:** 901-910

Telese F, Gamliel A, Skowronska-Krawczyk D, Garcia-Bassets I, Rosenfeld MG (2013) "Seq-ing" insights into the epigenetics of neuronal gene regulation. *Neuron* **77:** 606-623

Terns MP, Terns RM (2011) CRISPR-based adaptive immune systems. *Current opinion in microbiology* **14:** 321-327

Thomas KR, Capecchi MR (1987) Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* **51:** 503-512

Tomizawa S, Kobayashi H, Watanabe T, Andrews S, Hata K, Kelsey G, Sasaki H (2011) Dynamic stage-specific changes in imprinted differentially methylated regions during early mammalian development and prevalence of non-CpG methylation in oocytes. *Development* **138:** 811-820

Tong ZB, Hogberg H, Kuo D, Sakamuru S, Xia M, Smirnova L, Hartung T, Gerhold D (2016) Characterization of three human cell line models for high-throughput neuronal cytotoxicity screening. *Journal of applied toxicology : JAT*

Traynor J, Agarwa P, Lazzeroni L, Francke U (2002) Gene expression patterns vary in clonal cell cultures from Rett syndrome females with eight different MECP2 mutations. *BMC medical genetics* **3:** 1-15

Tsujimura K, Irie K, Nakashima H, Egashira Y, Fukao Y, Fujiwara M, Itoh M, Uesaka M, Imamura T, Nakahata Y, Yamashita Y, Abe T, Takamori S, Nakashima K (2015) miR-199a Links MeCP2 with mTOR Signaling and Its Dysregulation Leads to Rett Syndrome Phenotypes. *Cell reports*

Tudor M, Akbarian S, Chen RZ, Jaenisch R (2002) Transcriptional profiling of a mouse model for Rett syndrome reveals subtle transcriptional changes in the brain. *Proceedings of the National Academy of Sciences of the United States of America* **99:** 15536-15541

Tweedie S, Charlton J, Clark V, Bird A (1997) Methylation of Genomes and Genes at the Invertebrate-Vertebrate Boundary. *Molecular and cellular biology* **17:** 1469-1475

Urdinguio RG, Fernandez AF, Lopez-Nieva P, Rossi S, Huertas D, Kulis M, Liu CG, Croce CM, Calin GA, Esteller M (2010) Disrupted microRNA expression caused by Mecp2 loss in a mouse model of Rett syndrome. *Epigenetics : official journal of the DNA Methylation Society* **5:** 656-663

Urdinguio RG, Lopez-Serra L, Lopez-Nieva P, Alaminos M, Diaz-Uriarte R, Fernandez AF, Esteller M (2008) Mecp2-null mice provide new neuronal targets for Rett syndrome. *PloS one* **3:** e3669

Ure K, Lu H, Wang W, Ito-Ishida A, Wu Z, He L, Sztainberg Y, Chen W, Tang J, Zoghbi HY (2016) Restoration of Mecp2 expression in GABAergic neurons is sufficient to rescue multiple disease features in a mouse model of Rett Syndrome. *eLIFE*

Valinluck V, Tsai HH, Rogstad DK, Burdzy A, Bird A, Sowers LC (2004) Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2). *Nucleic acids research* **32:** 4100-4108

Van Esch H, Bauters M, Ignatius J, Jansen M, Raynaud M, Hollanders K, Lugtenberg D, Bienvenu T, Jensen LR, Gecz J, Moraine C, Marynen P, Fryns JP, Froyen G (2005) Duplication of the MECP2 region is a frequent cause of severe mental retardation and progressive neurological symptoms in males. *American journal of human genetics* **77:** 442-453

van Overbeek M, Capurso D, Carter MM, Thompson MS, Frias E, Russ C, Reece-Hoyes JS, Nye C, Gradia S, Vidal B, Zheng J, Hoffman GR, Fuller CK, May AP (2016) DNA Repair Profiling Reveals Nonrandom Outcomes at Cas9-Mediated Breaks. *Mol Cell* **63:** 633-646

Vandenberghe LH, Wang L, Somanathan S, Zhi Y, Figueredo J, Calcedo R, Sanmiguel J, Desai RA, Chen CS, Johnston J, Grant RL, Gao G, Wilson JM (2006) Heparin binding directs activation of T cells against adeno-associated virus serotype 2 capsid. *Nature medicine* **12:** 967-971

Varley KE, Gertz J, Bowling KM, Parker SL, Reddy TE, Pauli-Behn F, Cross MK, Williams BA, Stamatoyannopoulos JA, Crawford GE, Absher DM, Wold BJ, Myers RM (2013) Dynamic DNA methylation across diverse human cell lines and tissues. *Genome research* **23:** 555-567

Veeraragavan S, Wan YW, Connolly DR, Hamilton SM, Ward CS, Soriano S, Pitcher MR, McGraw CM, Huang SG, Green JR, Yuva LA, Liang AJ, Neul JL, Yasui DH, LaSalle JM, Liu Z, Paylor R, Samaco RC (2016) Loss of MeCP2 in the rat models regression, impaired sociability and transcriptional deficits of Rett syndrome. *Human molecular genetics* **25:** 3284-3302

Vives E, Brodin P, Lebleu B (1997) A Truncated HIV-1 Tat Protein Basic Domain Rapidly Translocates through the Plasma Membrane and Accumulates in the Cell Nucleus. *The Journal of biological chemistry* **272:** 16010-16017

Wakefield RID, Smith BO, Nan X, Free A, Soteriou A, Uhrin D, Bird AP, Barlow PN (1999) The Solution Structure of the Domain from MeCP2 that binds to Methylated DNA. *J Mol Biol*

Wang H, Yang H, Shivalila CS, Dawlaty MM, Cheng AW, Zhang F, Jaenisch R (2013a) One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* **153:** 910-918

Wang M, Zuris JA, Meng F, Rees H, Sun S, Deng P, Han Y, Gao X, Pouli D, Wu Q, Georgakoudi I, Liu DR, Xu Q (2016) Efficient delivery of genome-editing proteins using bioreducible lipid nanoparticles. *PNAS* **133:** 2868-2873

Wang X, Ghasri P, Amir M, Hwang B, Hou Y, Khalili M, Lin A, Keene D, Uitto J, Woodley DT, Chen M (2013b) Topical application of recombinant type VII collagen incorporates into the dermal-epidermal junction and promotes wound closure. *Molecular therapy : the journal of the American Society of Gene Therapy* **21:** 1335-1344

Weber M, Hellmann I, Stadler MB, Ramos L, Paabo S, Rebhan M, Schubeler D (2007) Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nature genetics* **39:** 457-466

Wender PA, Mitchell DJ, Pattabiraman K, Pelkey ET, Steinman L, Rothbard JB (2000) The design, synthesis, and evaluation of molecules that enable or enhance cellular uptake: Peptoid molecular transporters. *PNAS* **97:** 13003-13008

Wiedenheft B, Sternberg SH, Doudna JA (2012) RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **482:** 331-338

Will E, Klump H, Heffner N, Schwieger M, Schiedlmeier B, Ostertag W, Baum C, Stocking C (2002) Unmodified Cre recombinase crosses the membrane. *NAR* **30:** 1-6

Wong JJL, Gao D, Nguyen TV, Kwok C-T, van Geldermalsen M, Middleton R, Pinello N, Thoeng A, Nagarajah R, Holst J, Ritchie W, Rasko JEJ (2017) Intron retention is regulated by altered MeCP2-mediated splicing factor recruitment. *Nature communications* **8:** 15134

Wright AV, Nunez JK, Doudna JA (2016) Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. *Cell* **164:** 29-44

Wu H, Tao J, Chen PJ, Shahab A, Ge W, Hart RP, Ruan X, Ruan Y, Sun YE (2010) Genome-wide analysis reveals methyl-CpG-binding protein 2-dependent regulation of microRNAs in a mouse model of Rett syndrome. *Proceedings of the National Academy of Sciences of the United States of America* **107:** 18161-18166

Wu Y, Liang D, Wang Y, Bai M, Tang W, Bao S, Yan Z, Li D, Li J (2013) Correction of a Genetic Disease in Mouse via Use of CRISPR-Cas9. *Cell stem cell* **13:** 659-662

Xiang W, Schlachetzki JC, Helling S, Bussmann JC, Berlinghof M, Schaffer TE, Marcus K, Winkler J, Klucken J, Becker CM (2013) Oxidative stress-induced posttranslational modifications of alpha-synuclein: specific modification of alpha-synuclein by 4-hydroxy-2-nonenal increases dopaminergic toxicity. *Molecular and cellular neurosciences* **54:** 71-83

Xie W, Barr CL, Kim A, Yue F, Lee AY, Eubanks J, Dempster EL, Ren B (2012) Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. *Cell* **148:** 816-831

Yang C, van der Woerd MJ, Muthurajan UM, Hansen JC, Luger K (2011) Biophysical analysis and small-angle X-ray scattering-derived structures of MeCP2-nucleosome complexes. *Nucleic acids research* **39:** 4122-4135

Yang Y, Kucukkal TG, Li J, Alexov E, Cao W (2016) Binding Analysis of Methyl-CpG Binding Domain of MeCP2 and Rett Syndrome Mutations. *ACS chemical biology* **11:** 2706-2715

Yasui DH, Peddada S, Bieda MC, Vallero RO, Hogart A, Nagarajan RP, Thatcher KN, Farnham PJ, Lasalle JM (2007) Integrated epigenomic analyses of neuronal MeCP2 reveal a role for long-range interaction with active genes. *Proceedings of the National Academy of Sciences of the United States of America* **104:** 19416-19421

Yates A, Akanni W, Amode MR, Barrell D, Billis K, Carvalho-Silva D, Cummins C, Clapham P, Fitzgerald S, Gil L, Giron CG, Gordon L, Hourlier T, Hunt SE, Janacek SH, Johnson N, Juettemann T, Keenan S, Lavidas I, Martin FJ, Maurel T, McLaren W, Murphy DN, Nag R, Nuhn M, Parker A, Patricio M, Pignatelli M, Rahtz M, Riat HS, Sheppard D, Taylor K, Thormann A, Vullo A, Wilder SP, Zadissa A, Birney E, Harrow J, Muffato M, Perry E, Ruffier M, Spudich G, Trevanion SJ, Cunningham F, Aken BL, Zerbino DR, Flicek P (2016) Ensembl 2016. *Nucleic acids research* **44:** D710-716

Yazdani M, Deogracias Rn, Guy J, Port RA, Bird A, Barde Y-A (2012) Disease Modeling Using Embryonic Stem Cells: MeCP2 Regulates Nuclear Size and RNA Synthesis in Neurons. *Stem Cells*

Yokochi T, Robertson KD (2002) Preferential methylation of unmethylated DNA by Mammalian de novo DNA methyltransferase Dnmt3a. *The Journal of biological chemistry* **277:** 11735-11745

Yoon H-G, Chan DW, Huang Z-Q, Li J, Fondell JD, Qin J, Wong J (2003a) Purification and functional characterisation of the human N-CoR complex: The roles of HDAC3, TBL1 and TBLR1. *The EMBO journal* **22:** 1336-1346

Yoon H-G, Chan DW, Reynolds AB, Qin J, Wong J (2003b) N-CoR Mediates DNA Methylation-Dependent Repression through a Methyl CpG Binding Protein Kaiso. *Molecular Cell* **12:** 723-734

Young JI, Hong EP, Castle JC, Crespo-Barreto J, Bowman AB, Rose MF, Kang D, Richman R, Johnson JM, Berget S, Zoghbi HY (2005) Regulation of RNA splicing by the methylation-dependent transcriptional repressor methyl-CpG binding protein 2. *PNAS*

Yu F, Zingler N, Schumann G, Strätling WH (2001) Methyl-CpG-binding protein 2 represses LINE-1 expression and retrotransposition but not Alu transcription. *NAR* **29:** 4493–4501

Zahorakova D, Rosipal R, Hadac J, Zumrova A, Bzduch V, Misovicova N, Baxova A, Zeman J, Martasek P (2007) Mutation analysis of the MECP2 gene in patients of Slavic origin with Rett syndrome: novel mutations and polymorphisms. *Journal of human genetics* **52:** 342-348

Zhang W, Peterson M, Beyer B, Frankel WN, Zhang ZW (2014) Loss of MeCP2 from forebrain excitatory neurons leads to cortical hyperexcitation and seizures. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **34:** 2754-2763

Zhang X-Y, C.Supakar P, Khan R, C.Ehrlich K, Ehrlich M (1989) Related sites in human and herpesvirus DNA recognized by methylated DNA-binding protein from human placenta. *Nucleic acids research* **17:** 1459-1474

Zhang Y, Pak C, Han Y, Ahlenius H, Zhang Z, Chanda S, Marro S, Patzke C, Acuna C, Covy J, Xu W, Yang N, Danko T, Chen L, Wernig M, Sudhof TC (2013) Rapid single-step induction of functional neurons from human pluripotent stem cells. *Neuron* **78:** 785-798

Zhang Z-N, Freitas BC, Qian H, Lux J, Acab A, Trujillo CA, Herai RH, Nguyen Huu VA, Wen JH, Joshi-Barr S, Karpiak JV, Engler AJ, Fu X-D, Muotri AR, Almutairi A (2016) Layered hydrogels accelerate iPSC-derived neuronal maturation and reveal migration defects caused by MeCP2 dysfunction. *Proceedings of the National Academy of Sciences* **113:** 3185-3190

Zheng Y, Tesar DB, Benincosa L, Birnbock H, Boswell CA, Bumbaca D, Cowan KJ, Danilenko DM, Daugherty AL, Fielder PJ, Grimm HP, Joshi A, Justies N, Kolaitis G, Lewin-Koh N, Li J, McVay S, O'Mahony J, Otteneder M, Pantze M, Putnam WS, Qiu ZJ, Ruppel J, Singer T, Stauch O, Theil FP, Visich J, Yang J, Ying Y, Khawli LA, Richter WF (2012) Minipig as a potential translatable model for monoclonal antibody pharmacokinetics after intravenous and subcutaneous administration. *mAbs* **4:** 243-255

Zhou Z, Hong EJ, Cohen S, Zhao WN, Ho HY, Schmidt L, Chen WG, Lin Y, Savner E, Griffith EC, Hu L, Steen JA, Weitz CJ, Greenberg ME (2006) Brain-specific phosphorylation of MeCP2 regulates activity-dependent Bdnf transcription, dendritic growth, and spine maturation. *Neuron* **52:** 255-269

Ziller MJ, Muller F, Liao J, Zhang Y, Gu H, Bock C, Boyle P, Epstein CB, Bernstein BE, Lengauer T, Gnirke A, Meissner A (2011) Genomic distribution and inter-sample variation of non-CpG methylation across human cell types. *PLoS Genet* **7:** e1002389

Zuris JA, Thompson DB, Shu Y, Guilinger JP, Bessen JL, Hu JH, Maeder ML, Joung JK, Chen ZY, Liu DR (2015) Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. *Nature biotechnology* **33:** 73-80